# Forms of inquiry-based science instruction and their relations with learning outcomes: evidence from high and low-performing education systems

## Anindito Aditomo & Eckhard Klieme

View supplementary material ⬀

Published online: 25 Jan 2020.

Submit your article to this journal ⬀

View related articles ⬀

View Crossmark data ⬀

Routledge
Taylor & Francis Group

Check for updates

# Forms of inquiry-based science instruction and their relations with learning outcomes: evidence from high and low-performing education systems

Anindito Aditomo [a,b] and Eckhard Klieme [b]

[a]Faculty of Psychology, University of Surabaya, Surabaya, Indonesia; [b]Leibniz Institute for Research and Information in Education, Frankfurt am Main, Germany

**ABSTRACT**

Inquiry-based science instruction is widely advocated, but studies based on international large-scale assessments often show inquiry to be negatively associated with achievement. We re-examine this issue by examining whether the association between inquiry and learning depends upon the provision of teacher guidance. Participants were 151,721 students from 5089 schools from 10 highest and 10 lowest science performers in PISA 2015. Multi-group confirmatory factor analyses found that measurement invariance cannot be established, suggesting substantial regional variation in the pattern of inquiry-based instruction. Nonetheless, exploratory factor analyses indicated that at the conceptual level, many regions exhibit a pattern which contrasted between 'Guided inquiry' and 'Independent inquiry'. Results of structural equation modelling showed that inquiry is positively associated with outcomes when it incorporates teacher guidance, and negatively when it doesn't. However, the strength of the positive associations is stronger in regions where guided inquiry is measured with fewer items referring to student-centred activities. These findings are in line with current theories regarding the importance of scaffolding in learning from inquiry. This study suggests that it would be misguided to use PISA findings to support arguments to scale back inquiry and other constructivist approaches to teaching science.

Instructional approaches broadly described as 'inquiry-based' are considered to be essential for developing students' scientific literacy (Engeln, Mikelskis-seifert, & Euler, 2014). Studies based on international large-scale assessments (ILSA), however, often indicate inquiry to be associated with lower science achievement (Cairns & Areepattamannil, 2017; Chi, Liu, Wang, & Won Han, 2018; Grabau & Ma, 2017). Given the influence which ILSA can have on educational policy (Grek, 2009), such findings have raised concerns among advocates of inquiry-based instruction in science (Sjøberg, 2018; Zhang, 2016).

We re-examine this issue by taking a more nuanced look at the inquiry-based instruction measure in the Program for International Student Assessment (PISA) 2015. Unlike prior studies, we explicitly test the assumption of comparability (measurement invariance) across regions, and show that patterns of instruction are better characterised as being region-specific rather than universal. Building upon the variation of regional patterns, we also test the prediction that inquiry-based instruction is positively associated with learning outcomes when it involves some form of teacher guidance. In the following sections, we first summarise theoretical perspectives which link inquiry-based instruction and student learning. We then consider prior studies of inquiry-based science instruction based on ILSA and note areas of limitations which this study seeks to remedy.

## *Inquiry and science learning*

Inquiry-based instruction involves engaging students in formulating questions, collecting and analysing data, and reasoning and arguing about what the results mean (Barron & Darling-Hammond, 2008; Hmelo-silver, Duncan, & Chinn, 2007). Inquiry prompts active knowledge construction and thereby facilitates deeper learning (Barron & Darling-Hammond, 2008). It has become especially prominent in science education as part of the shift towards the 'practice turn' which recognises scientific practices as a central organising theme for teaching and learning (National Research Council, 2012).

The 'practice turn' is underpinned by socio-cultural theories with their central metaphor of learning as participation (Forman, 2018; Sfard, 1998). In this view, learning is the process of becoming a member of a community of practice. It is less about acquiring and having knowledge, and more about becoming able to participate in activities which are valued by a community, communicate using the language of that community, and act in ways which conform the community's norms. Science learning thus means participating in authentic scientific practices, i.e. constructing models/theories which explain some aspect of the natural world and arguing for their value and validity (Osborne, Simon, Christodoulou, & Howell-richardson, 2013; Windschitl, Thompson, & Braaten, 2008). Consequently, inquiry-based science instruction should weave together the conceptual, epistemic, and social dimensions of scientific practice (Duschl, 2008).

This conception of inquiry-based instruction highlights the non-cognitive dimensions of learning. Becoming competent participants of scientific practice involves changing beliefs about the nature of science, and thus good inquiry-based instruction should not only develop conceptual understanding, but also more mature epistemic beliefs (Sandoval, Greene, & Bråten, 2016). Such beliefs include, for instance, an understanding that scientific knowledge is subject to revision, and that knowledge is based on empirical evidence whose meaning is influenced by the models/theories which scientists employ (Duschl, 2007; Pluta, Chinn, & Duncan, 2011). In addition, inquiry may provide students more autonomy (e.g. in formulating questions and choosing how to address them) and opportunities for meaningful interactions and positive relationships. Thus, authentic inquiry has the potential to improve students' intrinsic motivation through the fulfilment of basic psychological needs (Ryan & Deci, 2000).

Critics charge that inquiry is unstructured and impose irrelevant cognitive which impede learning (Kirschner, Sweller, & Clark, 2006). However, inquiry-based instruction

does not necessarily be unstructured. Guided forms of inquiry incorporate various scaffolds to guide learners' meaning-making process. For instance, expert guidance can be embedded as 'just-in-time' mini-lectures; tasks can be sequenced to reduce cognitive load; and tools can be designed to model or make salient disciplinary strategies (Hmelo-silver et al., 2007). Indeed, experimental studies have shown that innovative inquiry-based interventions are superior to conventional science teaching (Furtak, Seidel, Iverson, & Briggs, 2016; Lazonder & Harmsen, 2016).

## Prior ILSA studies

ILSA such as PISA and TIMSS include measures designed to assess instructional practices, including inquiry-based ones. Accordingly, secondary analyses of ILSA data related to inquiry have been published. Some treated the PISA inquiry scale as a single index and found that higher frequencies of inquiry activities were related with lower science literacy (Cairns & Areepattamannil, 2017; Chi et al., 2018). A study using the TIMSS 2007 data also reported that 'student-oriented instruction', which reflect planning and conducting observations and investigations, was negatively associated with science achievement (Liou & Jessie Ho, 2018). Other studies found that different dimensions of the inquiry measure are related differently to outcomes. For example, science literacy was positively related with 'student investigation' activities, but negatively with 'hands-on' activities in the USA sample of PISA 2006 (Grabau & Ma, 2017). For the Qatar sample from the same data, however, both the hands-on and student investigation dimensions of inquiry were negatively related with science literacy (Areepattamannil, 2012).

Collectively these studies have contributed to the empirical base related to inquiry-oriented instruction as practiced in nationally representative schools in many countries/regions. A number of important limitations need to be noted, however. Methodologically, these studies assume that the same pattern of instructional practice exist across regions and can be measured using the same instrument. Given the possibility that instructional practices are region-specific, or that the measures function differently across the regions, measurement invariance needs to be explicitly tested (Wu, Li, & Zumbo, 2007). Conceptually, previous studies have not attempted to provide theoretically grounded explanations regarding the association between inquiry and learning. Without a theoretical account, findings such as differential relations between distinct dimensions of inquiry and learning outcomes are difficult to interpret. We argue that it is possible to propose and test a theoretical explanation regarding the relationship between inquiry and learning using ILSA data.

## Current study

Given the evidence from experimental studies about the efficacy of inquiry-based instruction, negative associations between inquiry and learning/achievement found in ILSA studies call for an explanation. One possibility is simply that the positive experimental evidence reflects the effects of innovative programmes in selected school/classroom settings, whereas findings from ILSA studies reflect inquiry activities as practiced in the 'regular' or typical school. This conjecture is supported by research which shows that successful enactment of inquiry-based curricula requires extensive training and support (Fitzgerald,

Danaia, & McKinnon, 2017; Fogleman, McNeill, & Krajcik, 2011), which is unlikely to be available for teachers in the 'average' school.

Another possible explanation is related to the level of guidance/structure. Teacher guidance can be seen as a form of structure necessary to facilitate learning, especially in complex activities such as scientific inquiry (Hmelo-silver et al., 2007; Schmidt, Loyens, & Paas, 2007). Without adequate guidance, learners may be overwhelmed by unessential features of the activity and fail to construct meaningful knowledge (Kirschner et al., 2006). Measures of inquiry-based instruction in PISA, specifically, include items which refer to student-independent activities (e.g. 'Students are asked to do an investigation to test ideas'), as well as those which refer to teacher guidance (e.g. 'The teacher explains how science ideas can be applied'). Thus, it may be possible to use PISA to test whether the association between inquiry and outcomes depends on teacher guidance.

These conjectures can be tested by comparing guided and unguided forms of inquiry. A recent study found that PISA's inquiry-based instruction scale form two separate dimensions, one reflecting teacher-guided interactive instruction and the other reflecting unguided inquiry (Lau & Lam, 2017). However, these authors utilised only 6 of the 9 available items, thereby further narrowing the scope of the construct. Furthermore, while the authors analysed data from high-performing regions, they did not examine whether the measurement model was statistically invariant/equivalent across the regions. Thus, the existence of alternative measurement models (reflecting different instructional patterns across the regions) cannot be ruled out.

We build upon and extend Lau and Lam's (2017) study in several ways. First, we expand the generality of the findings by analysing high and low-performing regions. It would be important to examine whether teacher guidance and inquiry can be effectively implemented by teachers in low-performing regions, where teacher competence is generally lower and school resources are more limited (Scheerens, 2001). Second, we explicitly test whether the single-factor structure (reflecting PISA's original design) and Lau and Lam's (2017) two-factor measurement models of inquiry-based instruction are statistically invariant across the selected regions. Third, we examine how inquiry relates to intrinsic motivation and epistemic beliefs as non-cognitive outcomes. We formulate the following research questions to structure our analysis and presentation of results:

(1) (a) Can the same forms of inquiry-based instructional practices be observed across high and low-performing regions, and (b) if not, what regional forms could be identified?

To answer this question, we tested the measurement invariance of a two-dimensional model which distinguishes between teacher-guided instruction and unguided inquiry based on Lau and Lam's (2017) study, and compared it to a unidimensional model of inquiry as intended by the PISA questionnaire designers (Müller, Prenzel, Seidel, Schiepe-tiska, & Kjærnsli, 2016).

(2) (a) How do the different forms of instruction relate to learning outcomes, and (b) how consistent is the relationship across various regions?

For this question, we employed structural equation modelling which incorporated instructional practices to predict three learning outcomes: science literacy, intrinsic

motivation, and epistemic beliefs. We expect learning outcomes to be related positively with forms of instruction which incorporate teacher guidance, and negatively with ones which do not.

## Method

### Sample and data

We examined nationally representative samples of 15-year-old students from 10 highest and 10 lowest performing regions in PISA 2015 (Table 1). For each region, PISA adopted a two-stage stratified sampling strategy in which randomly sampled schools and then 15-year-old students from each school (OECD, 2016). The total sample is composed of more than 150,000 students from 5089 schools. Participating students completed cognitive tests in science, math, and reading, as well as a background questionnaire which includes experiences of science instruction.

### Instructional practice measures

#### Inquiry-based instruction

We utilised nine items intended to measure inquiry-based instructional practices. The first two items refer explicitly to teacher guidance ('The teacher explains how science ideas can be applied' and 'The teacher clearly explains the relevance of science concepts to our lives'). One item ('Students are given the opportunity to explain their ideas') reflected a student-centred activity but does not refer to inquiry. The remainder (six items) explicitly refer to activities related to different aspects of inquiry (designing and conducting experiments, interpreting data, debating/arguing about science investigations, see Table 3).

**Table 1.** Regions and sample size.

| No | Region | N of schools | N of students |
|----|--------|--------------|---------------|
| *High-performing* | | | |
| 1 | B-S-J-G (China) | 268 | 9841 |
| 2 | Canada | 759 | 20,058 |
| 3 | Chinese Taipei | 214 | 7708 |
| 4 | Estonia | 206 | 5587 |
| 5 | Finland | 168 | 5882 |
| 6 | Hong Kong | 138 | 5359 |
| 7 | Japan | 198 | 6647 |
| 8 | Macao | 45 | 4476 |
| 9 | Singapore | 177 | 6115 |
| 10 | Vietnam | 188 | 5826 |
| *Low-performing* | | | |
| 11 | Algeria | 161 | 5519 |
| 12 | Brazil | 841 | 23,141 |
| 13 | Dominican Republic | 194 | 4740 |
| 14 | Indonesia | 236 | 6513 |
| 15 | Jordan | 250 | 7267 |
| 16 | Kosovo | 224 | 4826 |
| 17 | Lebanon | 270 | 4546 |
| 18 | Macedonia | 106 | 5324 |
| 19 | Peru | 281 | 6971 |
| 20 | Tunisia | 165 | 5375 |
| | TOTAL | 5089 | 151,721 |

### Transmissionist instruction

We also utilised four items measuring 'transmissionist instruction', i.e. a traditional teacher-centred mode in which content is delivered from teacher to students (example item: 'The teacher demonstrates an idea'). Incorporating this variable as a predictor in the models provides a benchmark to assist interpretation about the magnitude of associations between inquiry-based instruction and learning outcomes.

### Learning outcomes

PISA's science literacy score was used as the cognitive learning outcome variable in this study. The science literacy test measured students' ability to explain phenomena scientifically, evaluate and design scientific investigations, and interpret data and evidence. The test content is defined not by curriculum content, but rather by contexts and problems for which science concepts can be fruitfully applied. Due to time constraints, each student completed only part of the test and an IRT technique was used to derive 10 plausible values as estimates of students' science literacy.

Epistemic belief and intrinsic motivation (enjoyment of science) were examined as affective outcomes. Epistemic belief refers to personal views about the empirical basis and the evolving nature of scientific knowledge. Enjoyment of science refers to intrinsic motivation or the drive to learn science for the sake of the activity itself (Ryan & Deci, 2000).

### Covariates

Several variables which are known to correlate with academic achievement were used as co-variates: gender, immigrant status (whether one is an immigrant), mother tongue (whether is a native speaker of the test language), economic-social-cultural status (ESCS), and science self-efficacy. ESCS in PISA was a composite index which reflected parental education level and occupational status, cultural-educational resources at home, and overall family wealth. Science self-efficacy refers to subjective judgment about one's ability to perform actions and achieve certain goals related to science, e.g. to explain scientific phenomena and interpret data from scientific investigations. This variable was represented by an IRT-scaled score provided by the OECD.

### Analysis

Data analyses were performed using Mplus v.8 (Muthén & Muthén, 2017). All models were estimated using the robust maximum likelihood (MLR) estimator, which is robust against deviations from normal distributions and is also suitable for ordinal variables with at least four response categories (Scherer, Nilsen, & Jansen, 2016). Bias introduced by the two-stage stratified sampling was addressed by incorporating the final student weight variable (W_FSTUWT). Missing values were replaced using the full-information likelihood procedure in Mplus. The TYPE = COMPLEX setting in the ANALYSIS option in Mplus was used to correct for standard errors due to the clustered nature of the data (students nested within schools).[1]

### Dimensionality and measurement invariance

Two models were examined for their invariance across regions. The first is a model which combines all nine inquiry practice items in a single factor, representing the original PISA

design. The second is a model based on Lau and Lam (2017) which separates between an 'interactive application' factor (three items which did not refer to inquiry) and an 'inquiry' factor (six items which explicitly refer to inquiry activities). For each hypothesised structure, multi-group confirmatory factor analyses (MG-CFA) were implemented using the CONFIGURAL-METRIC-SCALAR setting within the MODEL option. The CONFIGURAL model constrained the factor structure but allowing item-factor loadings to vary. In the METRIC model, both factor structure and loadings were constrained to be equal across regions. Finally, in the SCALAR model, item intercepts were also constrained to be equal across regions (Muthén & Muthén, 2017).

The comparative fit index (CFI) and root mean square error approximation (RMSEA) were used to evaluate overall goodness-of-fit. Models were considered to have good fit if CFI was at least .95 and RMSEA not more than .08 (Rutkowski & Svetina, 2014). In the case of non-invariance, exploratory factor analyses (EFA) were performed separately for each region to identify instructional patterns in a bottom-up manner.

### Instruction and learning outcomes

To examine how instructional practices were related to learning, we added science literacy, enjoyment of science, and epistemic belief as outcome variables onto the measurement models identified in the previous step. All 10 science literacy plausible values were
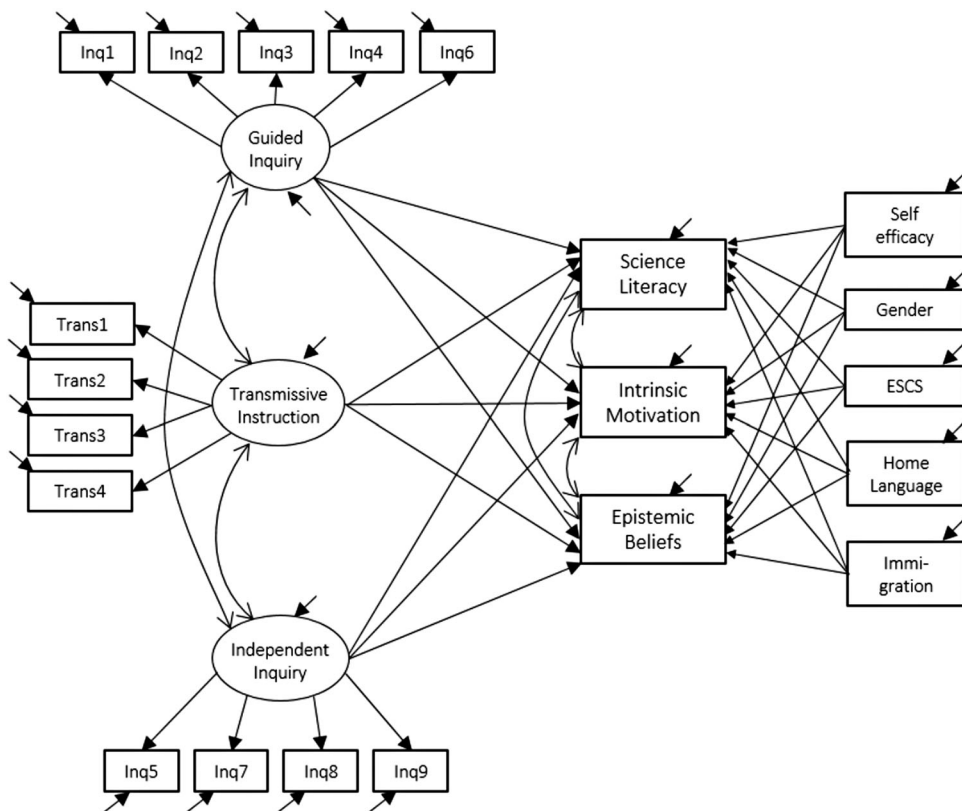


**Figure 1.** Measurement and structural model for Finland.

incorporated using the TYPE = IMPUTATION setting. In addition, ESCS, gender, immigrant status, language spoken at home, and science self-efficacy were included as co-variates. This amounted to structural equation modelling in which responses to the instructional practice items were included in the models as observed indicators which formed certain instructional practices as latent scores. As an illustration, Figure 1 displays the SEM model for Finland. The item content, see Table 3.

## Results

### Dimensionality and measurement invariance (RQ 1a)

Fit indices from the MG-CFA indicate that the 2-factor structure (teacher-guided instruction vs. unguided inquiry) outperformed the 1-factor structures (see Table 2). However, the poor fit of this model at the configural level suggests that the basic factor structure is not universal. Rather, there is substantial variation in the pattern of inquiry-based science instruction across the 20 regions. Uncovering these regional patterns of instruction requires an exploratory approach, the results of which are reported next.

### Regional forms and patterns of instruction (RQ 1b)

Given the lack of invariance, EFA were performed for each region to generate insights about regional forms and patterns of instruction. EFA results were then used to inform the construction and testing of a measurement model for each region. To test model fit, CFA was used for regions with low cross loadings (<0.2 on non-target factors), while Exploratory SEM (which allows items to cross load onto factors other than its target) was used for other regions. The factor loadings and fit indices are displayed in Tables 3–5.

Note that 'instructional form' refers to the formation of items which load together in a factor.[2] Meanwhile, we use 'instructional pattern' to refer to the combination of instructional forms which characterise the factor structure of a region. We first present findings regarding instructional *forms*, before commenting on instructional patterns, which were observed in the data.

Although no universal measurement model could be established, four instructional forms could be observed across the regions. Each instructional form is characterised by a certain combination of item loadings, and they could be arranged according to their degree of teacher guidance (Figure 2). While the specific items which compose an instructional form may vary across regions, their combination reflects the same conceptual meaning.

Thus, starting from the most student-centred, 'Independent Inquiry' is an instructional form composed only of items referring to inquiry activities, without any of the teacher

**Table 2.** Measurement invariance test results.

| Invariance level | | Configural | | Metric | | Scalar | |
|---|---|---|---|---|---|---|---|
| | | CFI | RMSEA | CFI | RMSEA | CFI | RMSEA |
| *Number of hypotesised factors* | 1 factor | 0.882 | 0.090 (0.090–0.091) | 0.868 | 0.084 (0.083–0.085) | 0.735 | 0.108 (0.108–0.109) |
| | 2 factors | 0.923 | 0.074 (0.073–0.075) | 0.910 | 0.072 (0.071–0.072) | 0.777 | 0.103 (0.102–0.104) |

**Table 3.** CFA measurement models for 'Guided vs Independent Inquiry' pattern.

| Items | Finland Guided Inq. | Finland Indep. Inq. | Macao Guided Inq. | Macao Indep Inq. | Singapore Guided Inq. | Singapore Indep. Inq. | Vietnam Guided Inq. | Vietnam Indep Inq. | Algeria Guided Inq. | Algeria Indep. Inq. | Kosovo Guided Inq. | Kosovo Indep Inq. | Lebanon Guided Inq. | Lebanon Indep. Inq. | Tunisia Guided Inq. | Tunisia Indep Inq. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| inq1. The teacher explains < school science > idea can be applied | 0.702 | | 0.565 | | 0.675 | | 0.469 | | 0.462 | | 0.478 | | 0.559 | | 0.552 | |
| inq2. The teacher clearly explains relevance < broad science > concepts to our lives. | 0.573 | | 0.514 | | 0.642 | | 0.507 | | 0.505 | | 0.495 | | 0.534 | | 0.604 | |
| inq3. Students are given opportunities to explain their ideas. | 0.466 | | 0.502 | | 0.557 | | 0.474 | | 0.433 | | 0.377 | | 0.468 | | 0.544 | |
| inq4. Students are asked to draw conclusions from an experiment they have conducted. | 0.696 | | 0.704 | | 0.655 | | 0.593 | | 0.486 | | | 0.813 | 0.587 | | 0.611 | |
| inq5. Students are required to argue about science questions. | | 0.729 | 0.754 | | | 0.687 | 0.626 | | 0.576 | | 0.613 | | 0.572 | | 0.691 | |
| inq6. Students spend time in the laboratory doing practical experiments. | 0.563 | | 0.691 | | 0.581 | | | 0.463 | 0.566 | | | 0.588 | | 0.414 | | 0.597 |
| inq7. Students are allowed to design their own experiments. | | 0.671 | | 0.717 | | 0.700 | | 0.654 | | 0.698 | | 0.737 | | 0.643 | | 0.680 |
| inq8. Students are asked to do an investigation to test ideas. | | 0.697 | | 0.758 | | 0.747 | | 0.663 | 0.578 | | 0.611 | | | 0.627 | | 0.704 |
| inq9. There is a class debate about investigations. | | 0.812 | | 0.802 | | 0.774 | | 0.649 | | 0.549 | 0.520 | | | 0.574 | | 0.698 |
| Fit indices | CFI = 0.960 RMSEA = 0.064 | | CFI = 0.962 RMSEA = 0.058 | | CFI = 0.954 RMSEA = 0.072 | | CFI = 0.951 RMSEA = 0.052 | | CFI = 0.956 RMSEA = 0.044 | | CFI = 0.950 RMSEA = 0.047 | | CFI = 0.961 RMSEA = 0.030 | | CFI = 0.952 RMSEA = 0.062 | |

**Table 4.** ESEM measurement models for 'Guided vs independent inquiry' pattern.

| Items | Canada | | B-S-J-G (China) | | Hong Kong | | Brazil | | Dominican | | Indonesia | | Jordan | | Peru | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Guided Inq. | Indep. Inq. | Guided Inq. | Indep Inq. | Guided Inq. | Indep. Inq. | Guided Inq. | Indep Inq. | Guided Inq. | Indep. Inq. | Guided Inq. | Indep Inq. | Guided Inq. | Indep. Inq. | Guided Inq. | Indep. Inq. 1 | Indep. Inq. 2 |
| inq1. The teacher explains < school science > idea can be applied | **0.923** | −0.139 | **0.855** | −0.097 | **0.97** | −0.213 | **0.464** | 0.275 | **0.586** | 0.039 | **0.322** | 0.175 | **0.633** | −0.027 | **0.75** | 0.066 | −0.119 |
| inq2. The teacher clearly explains relevance < broad science > concepts to our lives. | **0.616** | 0.134 | **0.519** | 0.238 | **0.86** | −0.082 | **0.916** | −0.081 | **0.748** | −0.155 | **0.369** | 0.086 | **0.784** | −0.206 | **0.775** | −0.204 | 0.143 |
| inq3. Students are given opportunities to explain their ideas. | **0.470** | 0.147 | **0.581** | 0.035 | **0.606** | 0.032 | 0.056 | **0.631** | **0.513** | 0.016 | **0.529** | 0.035 | **0.381** | 0.237 | **0.496** | 0.077 | 0.000 |
| inq4. Students are asked to draw conclusions from an experiment they have conducted. | **0.318** | **0.371** | 0.350 | **0.493** | **0.579** | 0.237 | −0.018 | **0.770** | **0.385** | **0.471** | **0.315** | **0.351** | 0.234 | **0.513** | 0.169 | **0.725** | 0.027 |
| inq5. Students are required to argue about science questions. | 0.096 | **0.658** | 0.217 | **0.613** | 0.169 | **0.654** | 0.133 | **0.669** | **0.519** | 0.228 | **0.806** | −0.122 | 0.24 | **0.368** | **0.481** | 0.277 | 0.021 |
| inq6. Students spend time in the laboratory doing practical experiments. | 0.081 | **0.637** | 0.112 | **0.659** | **0.488** | 0.286 | −0.215 | **0.654** | 0.068 | **0.526** | 0.009 | **0.456** | −0.118 | **0.867** | −0.004 | **0.519** | 0.188 |
| inq7. Students are allowed to design their own experiments. | −0.086 | **0.793** | −0.002 | **0.821** | 0.012 | **0.747** | −0.012 | **0.696** | 0.376 | 0.369 | 0.006 | **0.624** | 0.361 | 0.373 | −0.071 | 0.176 | **0.628** |
| inq8. Students are asked to do an investigation to test ideas. | 0.091 | **0.684** | −0.006 | **0.801** | **0.495** | 0.389 | 0.371 | 0.392 | **0.687** | −0.055 | 0.099 | **0.558** | 0.52 | 0.209 | **0.391** | −0.036 | **0.431** |
| inq9. There is a class debate about investigations. | −0.082 | **0.893** | −0.169 | **0.917** | 0.008 | **0.865** | 0.148 | **0.592** | **0.687** | 0.013 | −0.046 | **0.608** | 0.65 | 0.103 | 0.130 | −0.037 | **0.704** |
| *Fit indices* | *CFI = 0.968* *RMSEA = 0.051* | | *CFI = 0.979* *RMSEA = 0.056* | | *CFI = 0.974* *RMSEA = 0.061* | | *CFI = 0.965* *RMSEA = 0.051* | | *CFI = 0.964* *RMSEA = 0.054* | | *CFI = 0.972* *RMSEA = 0.038* | | *CFI = 0.971* *RMSEA = 0.051* | | *CFI = 0.983* *RMSEA = 0.055* | | |

**Table 5.** Measurement models for other instructional patterns.

| Items | Taipei | | Estonia | | Japan | | | Macedonia | |
|---|---|---|---|---|---|---|---|---|---|
| | Interactive Conceptual Instruction | Indep. Inquiry | Interactive Conceptual Instruction | Indep. Inquiry | Interactive Conceptual Instruction | Indep. Inquiry 1 | Indep. Inquiry 2 | Teacher-centred Conceptual Instruction | Indep. Inquiry |
| inq1. The teacher explains < school science > idea can be applied | 0.821 | | 0.740 | | 0.793 | | | **0.385** | 0.284 |
| inq2. The teacher clearly explains relevance < broad science > concepts to our lives. | 0.834 | | 0.732 | | 0.745 | | | **0.764** | −0.008 |
| inq3. Students are given opportunities to explain their ideas. | 0.532 | | 0.529 | | 0.488 | | | 0.22 | **0.306** |
| inq4. Students are asked to draw conclusions from an experiment they have conducted. | | 0.838 | | 0.686 | | | 0.892 | −0.016 | **0.695** |
| inq5. Students are required to argue about science questions. | | 0.831 | | 0.656 | | 0.723 | | 0.036 | **0.62** |
| inq6. Students spend time in the laboratory doing practical experiments. | | 0.744 | | 0.581 | | | 0.724 | −0.216 | **0.738** |
| inq7. Students are allowed to design their own experiments. | | 0.730 | | 0.682 | | 0.743 | | −0.049 | **0.697** |
| inq8. Students are asked to do an investigation to test ideas. | | 0.809 | | 0.647 | | 0.820 | | 0.226 | **0.482** |
| inq9. There is a class debate about investigations. | | 0.794 | | 0.723 | | 0.832 | | 0.104 | **0.566** |
| Fit indices | CFI = 0.977 RMSEA = 0.054 | | CFI = 0.957 RMSEA = 0.057 | | CFI = 0.950 RMSEA = 0.057 | | | CFI = 0.964 RMSEA = 0.035 | |
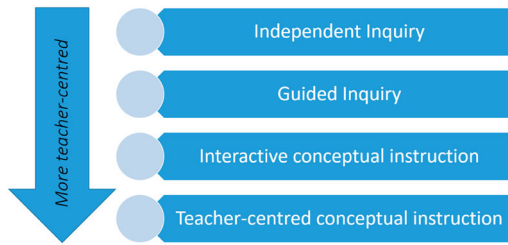
**Figure 2.** Forms of science instruction.

guidance items. Some variation of Independent Inquiry could be observed in all regions. Note that Independent Inquiry refers to instructional forms which centre on student-led inquiry activities but do not involve explicit conceptual teaching/content exposition from teachers. While it does not involve explicit conceptual teaching, we cannot rule out the possibility that some other forms of guidance are provided within the context of Independent Inquiry. For example, teachers may help structure students' group interactions or provide feedback regarding students' experimental designs.

The second instructional form, 'Guided Inquiry' combines the two teacher guidance items with at least one item referring to an inquiry activity. This form of inquiry was found in 16 of the 20 regions. The last two instructional forms involve some kind of teacher guidance, but without referring to any inquiry activities. Thus, 'Interactive Conceptual Instruction' is characterised by a combination of the two teacher guidance items with the one student-centred non-inquiry item ('Students are given opportunities to explain their ideas'). Interactive Conceptual Instruction was observed in Taipei, Estonia, and Japan. Last and the most teacher-centred, 'Teacher-centred Conceptual Instruction' is characterised simply by the two teacher guidance items. This instructional form was observed only in Macedonia.

Looking at the level of instructional pattern, 16 of the 20 regions contrasted between the two forms of inquiry, i.e. 'Independent Inquiry' and 'Guided Inquiry' (Tables 3 and 4). Meanwhile, the instructional patterns in the remaining four regions combined of 'Independent Inquiry' with either 'Interactive Conceptual Instruction' or 'Teacher-centred Conceptual Instruction'.

## Associations with outcomes (RQ 2a and 2b)

Results from structural models predicting learning outcomes are displayed in Tables 6–8 (note on $p$ values: ***$p < 0.001$; ** $p < 0.01$; *$p < 0.05$). We use the instructional pattern in 16 regions to test the hypothesis regarding the importance of teacher guidance for learning from inquiry. Regression results largely supported this hypothesis: guided inquiry was positively associated with science achievement in all 16 regions, with enjoyment in 15 regions, and with epistemic beliefs in 13 regions.

We use all 20 regions to test whether independent inquiry is negatively associated with learning outcomes. Again, the regression results by and large support this hypothesis: independent inquiry was found to be negatively associated with achievement in 18 regions, with enjoyment in 17 regions, and with epistemic beliefs also in 17 regions. The notable exception was Japan, where two forms of independent inquiry could be

**Table 6.** Standardised estimates (standard errors) from SEM models predicting literacy (cognitive achievement).

| Regions/predictors | Guided Inquiry | Independent Inquiry | Transmission. Instruction | Science self-efficacy | Econ-social-cultural status |
|---|---|---|---|---|---|
| Finland | 0.507*** (0.038) | −0.627*** (0.032) | −0.010 (0.021) | 0.225*** (0.014) | 0.208*** (0.013) |
| Macao[C] | 0.299*** (0.040) | −0.501*** (0.040) | 0.072*** (0.018) | 0.202*** (0.017) | 0.151*** (0.032) |
| Vietnam | 0.408*** (0.048) | −0.525*** (0.045) | 0.103*** (0.028) | 0.198*** (0.018) | 0.277*** (0.030) |
| Singapore | 0.495*** (0.048) | −0.546*** (0.042) | 0.044* (0.021) | 0.215*** (0.016) | 0.322*** (0.018) |
| Algeria | 0.291*** (0.058) | −0.418*** (0.054) | 0.048*** (0.023) | 0.016 (0.018) | 0.122*** (0.036) |
| Kosovo | 0.159** (0.060) | −0.366*** (0.055) | 0.190*** (0.019) | 0.052** (0.018) | 0.207*** (0.022) |
| Lebanon | 0.301*** (0.061) | −0.543*** (0.058) | 0.126*** (0.033) | 0.166*** (0.027) | 0.298*** (0.033) |
| Tunisia | 0.886*** (0.129) | −1.111*** (0.120) | 0.034 (0.032) | 0.081*** (0.017) | 0.300*** (0.030) |
| Canada | 0.321*** (0.019) | −0.482*** (0.015) | 0.035* (0.015) | 0.238*** (0.010) | 0.223*** (0.010) |
| Beijing | 0.595*** (0.030) | −0.556*** (0.023) | 0.040* (0.019) | 0.104*** (0.013) | 0.381*** (0.020) |
| Hong Kong | 0.326*** (0.032) | −0.492*** (0.031) | 0.109*** (0.022) | 0.149*** (0.016) | 0.183*** (0.019) |
| Brazil | 0.222*** (0.035) | −0.417*** (0.028) | 0.134*** (0.015) | 0.100*** (0.013) | 0.315*** (0.018) |
| Dominica | 0.149** (0.051) | −0.512*** (0.051) | 0.081** (0.028) | 0.001 (0.018) | 0.350*** (0.025) |
| Indonesia | 0.275*** (0.042) | −0.392*** (0.040) | 0.006 (0.020) | 0.021 (0.015) | 0.421*** (0.026) |
| Jordan | 0.150** (0.045) | −0.364*** (0.045) | 0.159*** (0.019) | 0.132*** (0.015) | 0.295*** (0.017) |
| Peru | 0.347*** (0.043) | −0.026 (0.039) and −0.569*** (0.035) | 0.086*** (0.020) | 0.045** (0.015) | 0.382*** (0.020) |
| **Regions/predictors** | Interactive conceptual inst. | Independent inquiry | Transmission. Instruction | Science self-efficacy | Econ.-social-cultural status |
| Estonia | 0.561*** (0.045) | −0.754*** (0.041) | −0.025 (0.019) | 0.154*** (0.018) | 0.219*** (0.018) |
| Taipei | 0.485*** (0.025) | −0.541*** (0.027) | 0.045** (0.013) | 0.189*** (0.013) | 0.261*** (0.016) |
| Macedonia | 0.367*** (0.082) | −0.446*** (0.073) | 0.073*** (0.021) | 0.153*** (0.024) | 0.232*** (0.026) |
| Japan | 0.130*** (0.031) | −0.483*** (0.032) and 0.254*** (0.038) | 0.082*** (0.021) | 0.182*** (0.012) | 0.244*** (0.015) |

**Table 7.** Standardised estimates (standard errors) from SEM models predicting enjoyment.

| Regions/predictors | Guided Inquiry | Independent Inquiry | Transmission. Instruction | Science self-efficacy | Econ-social-cultural status |
|---|---|---|---|---|---|
| Finland | 0.252*** (0.034) | −0.125*** (0.032) | 0.157*** (0.022) | 0.303*** (0.016) | 0.076*** (0.014) |
| Macao | 0.137*** (0.038) | −0.088** (0.029) | 0.121*** (0.021) | 0.318*** (0.019) | 0.027 (0.019) |
| Vietnam | 0.028 (0.042) | −0.060 (0.038) | 0.193*** (0.031) | 0.203*** (0.019) | 0.022 (0.020) |
| Singapore | 0.387*** (0.051) | −0.245*** (0.039) | 0.094*** (0.023) | 0.319*** (0.019) | 0.051*** (0.014) |
| Algeria | 0.286*** (0.045) | −0.234*** (0.048) | 0.214*** (0.022) | 0.077*** (0.020) | −0.026 (0.015) |
| Kosovo | 0.169*** (0.040) | −0.081** (0.037) | 0.173*** (0.019) | 0.042 (0.023) | 0.018 (0.018) |
| Lebanon | 0.213*** (0.053) | −0.189*** (0.051) | 0.278*** (0.031) | 0.174*** (0.027) | 0.075** (0.024) |
| Tunisia | 0.563*** (0.108) | −0.498*** (0.102) | 0.167*** (0.030) | 0.107*** (0.023) | 0.019 (0.015) |
| Canada | 0.276*** (0.022) | −0.161*** (0.017) | 0.115*** (0.016) | 0.326*** (0.013) | 0.066*** (0.010) |
| Beijing | 0.204*** (0.029) | −0.026 (0.026) | 0.122*** (0.018) | 0.252*** (0.018) | 0.086*** (0.015) |
| Hong Kong | 0.241*** (0.032) | −0.144*** (0.028) | 0.179*** (0.030) | 0.342*** (0.020) | 0.026 (0.015) |
| Brazil | 0.227*** (0.025) | −0.061* (0.024) | 0.158*** (0.016) | 0.188*** (0.016) | 0.074*** (0.013) |
| Dominica | 0.076* (0.032) | −0.065 (0.039) | 0.233*** (0.024) | 0.101*** (0.026) | −0.029 (0.019) |
| Indonesia | 0.203*** (0.038) | −0.104** (0.035) | 0.148*** (0.023) | 0.110*** (0.016) | −0.019 (0.019) |
| Jordan | 0.204*** (0.036) | −0.210*** (0.034) | 0.288*** (0.022) | 0.208*** (0.021) | 0.006 (0.017) |
| Peru | 0.150*** (0.036) | −0.080** (0.031) and −0.062* (0.029) | 0.162*** (0.023) | 0.178*** (0.018) | −0.017 (0.016) |
| **Regions/predictors** | Interactive conceptual inst. | Independent inquiry | Transmission. Instruction | Science self-efficacy | Econ.-social-cultural status |
| Estonia | 0.343*** (0.037) | −0.280*** (0.039) | 0.118*** (0.018) | 0.202*** (0.022) | 0.084*** (0.017) |
| Taipei | 0.135*** (0.021) | −0.006 (0.019) | 0.083*** (0.013) | 0.330*** (0.012) | 0.075*** (0.012) |
| Macedonia | 0.292*** (0.055) | −0.217*** (0.055) | 0.204*** (0.018) | 0.098*** (0.027) | −0.017 (0.021) |
| Japan | 0.244*** (0.027) | −0.149*** (0.024) and 0.014 (0.025) | 0.102*** (0.020) | 0.314*** (0.015) | 0.071*** (0.014) |

**Table 8.** Standardised estimates (standard errors) from SEM models predicting epistemic beliefs.

| Regions/predictors | Guided Inquiry | Independent Inquiry | Transmission. Instruction | Science self-efficacy | Econ-social-cultural status |
|---|---|---|---|---|---|
| *Finland* | 0.312*** (0.042) | −0.317*** (0.039) | 0.050* (0.025) | 0.176*** (0.020) | 0.133*** (0.014) |
| *Macao* | 0.126*** (0.036) | −0.129*** (0.029) | 0.144*** (0.022) | 0.181*** (0.018) | 0.050** (0.019) |
| *Vietnam* | 0.040 (0.043) | −0.151*** (0.040) | 0.231*** (0.026) | 0.113*** (0.022) | 0.139*** (0.018) |
| *Singapore* | 0.410*** (0.049) | −0.347*** (0.041) | 0.064** (0.023) | 0.202*** (0.020) | 0.073*** (0.015) |
| *Algeria* | 0.086 (0.044) | −0.108* (0.042) | 0.165*** (0.023) | 0.022 (0.028) | 0.031 (0.016) |
| *Kosovo* | 0.193*** (0.039) | −0.170*** (0.036) | 0.137*** (0.022) | −0.041 (0.026) | 0.076*** (0.020) |
| *Lebanon* | 0.337*** (0.060) | −0.375*** (0.059) | 0.154*** (0.033) | 0.116*** (0.034) | 0.098*** (0.031) |
| *Tunisia* | 0.320*** (0.095) | −0.343*** (0.089) | 0.134*** (0.029) | 0.054* (0.023) | 0.078*** (0.019) |
| *Canada* | 0.230*** (0.024) | −0.231*** (0.018) | 0.097*** (0.018) | 0.175*** (0.013) | 0.105*** (0.010) |
| *Beijing* | 0.220*** (0.029) | −0.146*** (0.025) | 0.074*** (0.018) | 0.191*** (0.022) | 0.141*** (0.015) |
| *Hong Kong* | 0.277*** (0.035) | −0.257*** (0.033) | 0.152*** (0.025) | 0.193*** (0.025) | 0.074*** (0.017) |
| *Brazil* | 0.160*** (0.026) | −0.162*** (0.025) | 0.169*** (0.018) | 0.097*** (0.016) | 0.101*** (0.014) |
| *Dominica* | 0.066* (0.033) | −0.164*** (0.038) | 0.150*** (0.030) | 0.015 (0.021) | 0.045* (0.020) |
| *Indonesia* | 0.031 (0.034) | −0.007 (0.034) | 0.102*** (0.023) | 0.029 (0.021) | 0.103*** (0.017) |
| *Jordan* | 0.200*** (0.036) | −0.255*** (0.032) | 0.223*** (0.023) | 0.125*** (0.021) | 0.088*** (0.016) |
| *Peru* | 0.136*** (0.038) | −0.015 (0.028) and −0.184*** (0.032) | 0.134*** (0.021) | 0.057** (0.019) | 0.135*** (0.014) |
| **Regions/predictors** | **Interactive conceptual inst.** | **Independent inquiry** | **Transmission. Instruction** | **Science self-efficacy** | **Econ.-social-cultural status** |
| *Estonia* | 0.393*** (0.037) | −0.423*** (0.039) | 0.074** (0.021) | 0.080*** (0.023) | 0.094*** (0.017) |
| *Taipei* | 0.287*** (0.023) | −0.246*** (0.022) | 0.078*** (0.015) | 0.186*** (0.016) | 0.118*** (0.013) |
| *Macedonia* | 0.241*** (0.063) | −0.237*** (0.058) | 0.192*** (0.020) | 0.040 (0.025) | 0.114*** (0.018) |
| *Japan* | 0.182*** (0.030) | −0.219*** (0.027) and 0.051* (0.025) | 0.144*** (0.021) | 0.214*** (0.019) | 0.129*** (0.015) |

observed, one of which was positively associated with achievement and epistemic beliefs.

Additional analysis[3] found that the positive associations between Guided Inquiry and achievement (science literacy) were weaker in regions were Guided Inquiry was measured using more inquiry-specific items. In other words, Guided Inquiry which includes more student-centred inquiry activities seems to be less strongly associated with achievement, compared to Guided Inquiry which includes fewer inquiry activities. To illustrate, in the Dominican Republic, where Guided Inquiry included five inquiry items, the effect size (predicting literacy) was 0.149. Meanwhile, the corresponding effect sizes were 0.291 in Algeria (where Guided Inquiry included 4 inquiry-specific items), 0.299 in Macao (with 3 inquiry-specific items), 0.408 in Vietnam (with 2 inquiry-specific items), and 0.495 in Singapore (with 1 inquiry-specific item). Further analysis at the item level showed that all inquiry items were associated with higher science literacy in all countries except Vietnam (where the correlations were positive but not significant).

## Discussion

The current study takes a more nuanced look at the measurement of inquiry-based instruction and how it relates to learning outcomes in the highest and lowest performing regions of PISA 2015. Prior studies have assumed, without explicitly testing, that the structure of inquiry-based instruction is equivalent across regions. Our examination found little support for this assumption of measurement invariance. That is, the nine items designed to assess inquiry-based instruction in PISA do not form the same universal structure. Rather, the analysis revealed regional patterns of instruction. Nonetheless, at a higher level of abstraction, our analysis also suggests that in many regions, a distinction between 'Guided Inquiry' and 'Independent Inquiry' can be found. Both forms of instruction involve the use of inquiry activities. The difference between them is that Guided Inquiry combines inquiry activities with teachers' explanations about how science concepts can be applied.

The contrast between Guided and Independent Inquiry allowed us to test the conjecture that, when coupled with teacher guidance, inquiry can be associated with better learning outcomes. Results of the structural equation modelling provide strong support for this conjecture for all types of outcomes examined. Guided Inquiry was positively associated with scores on science achievement test in all the 16 regions where this form of instruction was observed. Positive associations between Guided Inquiry and affective outcomes (intrinsic motivation and epistemic beliefs) could also be found in the majority of the 16 regions.

Conversely, Independent Inquiry was found to be negatively associated with learning outcomes in 19 of the 20 regions where this form of instruction could be observed. As previously noted, the exception was Japan, where one form of Independent Inquiry was positively associated with science achievement and epistemic belief. This suggests that, in some contexts, inquiry can be effective even without teacher guidance.

This study brings PISA-based findings in line with mainstream theories of learning and empirical findings in science education (Furtak et al., 2016; Lazonder & Harmsen, 2016). As elaborated in the Introduction, cognitive and socio-constructivist theories emphasise the importance of scaffolding for learning from inquiry (Hmelo-silver et al., 2007).

Accordingly, this study found that inquiry activities were associated positively with learning outcomes when they are coupled with teachers' conceptual guidance. Further lending support for this conclusion is the observation that the positive association between Guided Inquiry and achievement seemed to be weaker in regions where Guided Inquiry was more student-centred (measured using more items which reflect student-centred inquiry activities).

More importantly, our findings not only confirm theoretical predictions. They also serve as evidence that teachers teaching in the 'average' or typical school are able to provide guidance which makes inquiry meaningful and effective. Moreover, this was true not only for high-performing education systems, but also for low-performing ones. In other words, Guided Inquiry seems to be effective (and more so than traditional instruction) even when implemented in schools with more limited resources and in education systems where teacher and teaching quality are generally poor (Aslam et al., 2016; Scheerens, 2001).

A critical reader might question whether the statistically significant associations found in this study are also practically meaningful. This question needs to be addressed especially because this study utilised large sample sizes which increase the possibility of Type I error ('false positives'). The meaning of effect sizes is difficult to judge in absolute terms. One way to judge the practical significance of this study's findings is by comparing it to effect sizes found in prior studies. The associations between inquiry-based instruction and learning outcomes found in most regions in this study are comparable to the effects sizes summarised in meta-analyses of educational effectiveness studies (Kyriakides, Christoforou, & Charalambous, 2013; Scheerens, Luyten, Steen, & Luyten-De Thouars, 2007). In this metric, the effects observed in the present study can be considered as moderate.

Another way to gauge magnitude of the observed effects is through comparisons with the effect sizes other predictors of learning outcomes. Using this metric, effect sizes of Guided Inquiry on achievement are larger and more consistently positive than the effect sizes of Transmissionist Instruction, especially when looking at cognitive outcome. In addition, the effect sizes of Guided Inquiry on achievement are also often larger than, or at least comparable to, the effects of science self-efficacy as well as family economic-socio-cultural background. Thus, we argue that the magnitude of associations observed in the current study can be considered meaningful.

These findings are significant because ILSA of learning such as PISA can exert significant influence on educational policy (Berliner, 2015; Grek, 2009). While its cross-sectional design prevents causal inferences to be made, findings from ILSA are perceived to have strong external validity because they are based on nationally representative samples of schools and students. With regard to science teaching, analysis based on ILSA data often shows inquiry to be negatively associated with science achievement (Cairns & Areepattamannil, 2017; Chi et al., 2018). This finding can lead to the suggestion that teachers who teach in the 'typical' school may not have the capacity or support required to implement inquiry effectively. Thus, some have voiced concerns that the desire to climb the 'PISA ladder' may prompt policy makers to discourage the use of inquiry-based instruction (Sjøberg, 2018). The current study, however, suggests that prior negative associations between inquiry-based instruction and learning were likely due to the conflation between guided and unguided forms of inquiry. In accordance with the mainstream

view in science education, this study finds that when combined with teacher guidance, inquiry is positively associated with cognitive and affective learning outcomes. Furthermore, in almost all regions the positive effects of guided inquiry on learning were found to be larger than transmissionist instruction. Thus, it would be misguided to use PISA findings to support arguments which favour explicit/direct forms of instruction over constructivist approaches such as inquiry.

Another point worth discussing is that the instructional forms observed in this study may not reflect the kind of authentic inquiry advocated by science educators (Chinn & Malhotra, 2002). In none of the regions did all nine items intended to measure inquiry-based instruction form a single dimension. In other words, according to the students in our sample, science teachers tend to employ only limited sets of inquiry activities. In a sense, this finding is unsurprising. Interweaving the empirical, epistemic, social, and conceptual dimensions to enact authentic inquiry is no easy feat (Harris & Rooks, 2010). Nonetheless, from a practice point of view, the findings here suggest that it doesn't really matter which aspect of inquiry are implemented. It matters little whether a teacher asks students to design and conduct lab-based experiments, or another provides empirical data for students to discuss and debate. Rather, what matters for student learning is whether teachers are actively involved to help students make sense of and conceptualise their inquiry activity.

On the issue of measurement, another limitation of this study stems from the fact that PISA's inquiry scale was not designed to measure guidance in the context of inquiry. More specifically, while the scale included items reflecting teacher conceptual guidance, it does not measure other forms of guidance or structure which may be incorporated to scaffold students' inquiry. Consequently, the forms of 'Independent Inquiry' identified in this study need to be interpreted with caution, as they may include forms of guidance which are not measured by the scale. This points to the need to design instruments which explicitly assess the use of scaffolding and guidance in the context of inquiry-based instruction.

Last, we note a general limitation of findings based on ILSA which stems from the cross-sectional nature of the data. In examining the relations between instruction and learning outcomes, it is difficult to ascertain the direction of causality. It is possible that teachers tend to refrain from providing conceptual guidance to students who – at the outset of instruction – exhibit low interest, efficacy, motivation, and/or achievement. In other words, students' initial motivation and achievement maybe the driving force for which type of instruction teachers employ. In this study, we have attempted to address this problem by including important determinants of achievement as co-variates, including students' science self-efficacy and socio-economic background. While this may partially mitigate the issue, we recognise that no strong causal inference can be made with regard to inquiry-based instruction and learning outcomes. Future studies should strive to include measures of prior achievement, ideally within a longitudinal design, to address this issue.

## Conclusions and implications

In closing, we conclude that suggestions to discourage science teachers from utilising inquiry activities seem to be misguided. When examined in more detail, ILSA data

yield findings consistent with mainstream theory and experimental studies. In this study, we show that student-centred inquiry activities, in and of themselves, are not the culprit of low motivation or achievement in science classes. On the contrary, inquiry activities tend to be associated with higher outcomes when coupled with teachers' active involvement to help students make sense or conceptualise their experiences. Significantly, this study suggests that the 'average' teacher teaching in the 'average' school is capable of providing the type of conceptual guidance needed to facilitate productive science learning.

For future research, one implication arising from our findings is that researchers should pay careful attention to issues of measurement invariance when examining instructional practices in ILSA data. Another implication is that future ILSA would benefit from developing measures specifically designed to assess the quality of teacher guidance in inquiry. With regard to practical implications, teacher-guided inquiry activities seem to be an essential component of instruction when the goal is to develop students' scientific literacy. Even lower performing education systems would benefit from encouraging teachers to couple one or another inquiry activity with conceptual explanations. Also, to the extent that some teachers view student-centred teaching as equating to letting students on their own course without providing structure and guidance, policy documents and teacher training should counter such misconceptions.

## Notes

1. The research questions in this study deal with relationships of variables at a single (student) level, and thus preclude the need for multilevel modelling (Stapleton, McNeish, & Yang, 2016). The TYPE=COMPLEX approach was preferred because it handles the clustered/ non-independent observations while requiring substantially less computational time compared to multilevel modelling (Muthen & Satorra, 1995).
2. As further evidence of the validity of the distinction between Guided and Independent Inquiry, we examined the correlations between both inquiry forms and three teaching variables: Transmissionist Instruction, Adaptive Instruction, and Emotional Support. Both forms of inquiry were positively correlated with the other teaching variables in the vast majority of the regions. More importantly, correlations were stronger with Guided Inquiry compared to Independent Inquiry in all regions. We argue that this is strong evidence supporting our interpretation of the conceptual distinction between the two forms of inquiry. On average, Guided Inquiry correlated with Transmissionist Instruction at 0.38 (range −0.02 to 0.59), with Adaptive Instruction at 0.46 (range 0.32 to 0.60), and with Emotion Support at 0.59 (range 0.47 to 0.73). Meanwhile, on average Independent Inquiry correlated with Transmissionist Instruction at 0.21 (range −0.16 to 0.49), with Adaptive Instruction at 0.27 (range 0.07 to 0.41), and with Emotional Support at 0.34 (range 0.17 to 0.50). See the Online Appendix for complete results.
3. We thank an anonymous reviewer for pointing this out by analysing the correlation between number of inquiry-specific items and Guided Inquiry effect size (across regions).

## Disclosure statement

## Funding

## ORCID

*Anindito Aditomo* 🔟 http://orcid.org/0000-0003-3711-3773
*Eckhard Klieme* 🔟 http://orcid.org/0000-0003-0728-4950

## References

Areepattamannil, S. (2012). Effects of inquiry-based science instruction on science achievement and interest in science: Evidence from Qatar. *The Journal of Educational Research*, *105*(2), 134–146. doi:10.1080/00220671.2010.533717

Aslam, M., Rawal, S., Kingdon, G., Moon, B., Banerji, R., Das, S., … Sharma, S. K. (2016). *Reforms to increase teacher effectiveness in developing countries*. London: EPPI-Centre, Social Science Research Unit, UCL Institute of Education, University College London.

Barron, B., & Darling-Hammond, L. (2008). *Teaching for meaningful learning: A review of research on inquiry-based and cooperative learning*. San Rafael, CA: George Lucas Educational Foundation.

Berliner, D. C. (2015). The many facets of PISA. *Teachers College Record*, *117*(1), 1–20.

Cairns, D., & Areepattamannil, S. (2017). Exploring the relations of inquiry-based teaching to science achievement and dispositions in 54 countries. *Research in Science Education*, 1–23. doi:10.1007/s11165-017-9639-x

Chi, S., Liu, X., Wang, Z., & Won Han, S. (2018). Moderation of the effects of scientific inquiry activities on low SES students' PISA 2015 science achievement by school teacher support and disciplinary climate in science classroom across gender. *International Journal of Science Education*, 1–21. doi:10.1080/09500693.2018.1476742

Chinn, C. A., & Malhotra, B. A. (2002). Epistemologically authentic inquiry in schools: A theoretical framework for evaluating inquiry tasks. *Science Education*, *86*, 175–218. doi:10.1002/sce.10001

Duschl, R. A. (2007). Quality argumentation and epistemic criteria. In S. Erduran & M. P. Jimenez-Aleixandre (Eds.), *Argumentation in science education: Perspectives from classroom-based research* (pp. 159–175). Dordrecht: Springer.

Duschl, R. A. (2008). Science education in three-part harmony: Balancing conceptual, epistemic, and social learning goals. *Review of Research in Education*, *32*, 268–291. doi:10.3102/0091732X07309371

Engeln, K., Mikelskis-seifert, S., & Euler, M. (2014). Inquiry-based mathematics and science education across Europe: A synopsis of various approaches and their potentials. In C. Bruguière (Ed.), *Topics and Trends in current science education: 9th ESERA conference selected contributions* (pp. 229–242). Dordrecht: Springer Science. doi:10.1007/978-94-007-7281-6

Fitzgerald, M., Danaia, L., & McKinnon, D. H. (2017). Barriers inhibiting inquiry-based science teaching and potential solutions: Perceptions of positively inclined early adopters. *Research in Science Education*, 1–24. doi:10.1007/s11165-017-9623-5

Fogleman, J., McNeill, K. L., & Krajcik, J. (2011). Examining the effect of teachers' adaptations of a middle school science inquiry-oriented curriculum unit on student learning. *Journal of Research in Science Teaching*, *48*(2), 149–169. doi:10.1002/tea.20399

Forman, E. A. (2018). The practice turn in learning theory and science education. In David W. Kritt (Ed.), *Constructivist education in an age of accountability* (pp. 97–104). Cham, Switzerland: Palgrave Macmillan. doi:10.1007/978-3-319-66050-9_5

Furtak, E. M., Seidel, T., Iverson, H., & Briggs, D. C. (2016). Experimental and quasi-experimental studies of inquiry-based science teaching: A meta-analysis. *Review of Educational Research*, *82*(3), 300–329. doi:10.3102/0034654312457206

Grabau, L. J., & Ma, X. (2017). Science engagement and science achievement in the context of science instruction: A multilevel analysis of U.S. students and schools. *International Journal of Science Education*, *39*(8), 1045–1068. doi:10.1080/09500693.2017.1313468

Grek, S. (2009). Governing by numbers: The PISA "effect" in Europe. *Journal of Education Policy*, *24*(1), 23–37. doi:10.1080/02680930802412669

Harris, C. J., & Rooks, D. L. (2010). Managing inquiry-based science: Challenges in enacting complex science instruction in elementary and middle school classrooms. *Journal of Science Teacher Education*, 21(2), 227–240. doi:10.1007/s10972-009-9172-5

Hmelo-silver, C. E., Duncan, R. G., & Chinn, C. A. (2007). Scaffolding and achievement in problem-based and inquiry learning: A response to Kirschner, Sweller, and Clark (2006). *Educational Psychologist*, 42(2), 99–107.

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, 41(2), 75–86. doi:10.1207/s15326985ep4102

Kyriakides, L., Christoforou, C., & Charalambous, C. Y. (2013). What matters for student learning outcomes: A meta-analysis of studies exploring factors of effective teaching. *Teaching and Teacher Education*, 36, 143–152. doi:10.1016/j.tate.2013.07.010

Lau, K., & Lam, T. Y. (2017). Instructional practices and science performance of 10 top-performing regions in PISA 2015. *International Journal of Science Education*, 1–22. doi:10.1080/09500693.2017.1387947

Lazonder, A. W., & Harmsen, R. (2016). Meta-analysis of inquiry-based learning: Effects of guidance. *Review of Educational Research*, 86(3), 681–718. doi:10.3102/0034654315627366

Liou, P.-Y., & Jessie Ho, H.-N. (2018). Relationships among instructional practices, students' motivational beliefs and science achievement in Taiwan using hierarchical linear modelling. *Research Papers in Education*, 33(1), 73. doi:10.1080/02671522.2016.1236832

Müller, K., Prenzel, M., Seidel, T., Schiepe-tiska, A., & Kjærnsli, M. (2016). Science teaching and learning in schools: Theoretical and empirical foundations for investigating classroom-level processes. In S. Kuger, E. Klieme, N. Jude, & D. Kaplan (Eds.), *Assessing contexts of learning: An international Perspective* (pp. 423–446). Cham, Switzerland: Springer International. doi:10.1007/978-3-319-45357-6

Muthén, L. K., & Muthén, B. O. (2017). *Mplus user's guide* (8th ed.). Los Angeles, CA: Muthén & Muthén.

Muthen, B. O., & Satorra, A. (1995). Complex sample data in structural equation modeling. *Sociological Methodology*, 25, 267–316. doi:10.1016/j.acthis.2013.11.004

National Research Council. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. Washington, DC: The National Academies.

OECD. (2016). *PISA 2015 results (Volume I): Excellence and equity in education (Vol. I)*. Paris: Author.

Osborne, J., Simon, S., Christodoulou, A., & Howell-richardson, C. (2013). Learning to argue: A study of four schools and their attempt to develop the use of argumentation as a common instructional practice and its impact on students. *Journal of Research in Science Teaching*, 50(3), 315–347. doi:10.1002/tea.21073

Pluta, W. J., Chinn, C. A., & Duncan, R. G. (2011). Learners' epistemic criteria for good scientific models. *Journal of Research in Science Teaching*, 48(5), 486–511.

Rutkowski, L., & Svetina, D. (2014). Assessing the hypothesis of measurement invariance in the context of large-scale international surveys. *Educational and Psychological Measurement*, 74(1), 31–57. doi:10.1177/0013164413498257

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55(1), 68–78. doi:10.1037/0003-066X.5501.68

Sandoval, W. A., Greene, J. A., & Bråten, I. (2016). Understanding and promoting thinking about knowledge origins, issues, and future directions of research on epistemic cognition. *Review of Research in Education*, 40(1), 457–496.

Scheerens, J. (2001). Monitoring school effectiveness in developing countries. *School Effectiveness and School Improvement*, 12(4), 359–384. doi:10.1076/sesi.12.4.359.3447

Scheerens, J., Luyten, H., Steen, R., & Luyten-De Thouars, Y. (2007). *Review and meta-analyses of school and teaching effectiveness*. Retrieved from https://www.utwente.nl/nl/bms/omd/Medewerkers/artikelen/scheerens/rev-meta-analysis2007.PDF

Scherer, R., Nilsen, T., & Jansen, M. (2016). Evaluating individual students' perceptions of instructional quality: An investigation of their factor structure, measurement invariance, and relations to educational outcomes. *Frontiers in Psychology*, *7*(February), 1–16. doi:10.3389/fpsyg.2016.00110

Schmidt, H. G., Loyens, S. M. M., & Paas, F. (2007). Problem-based learning is compatible with human cognitive architecture: Commentary on Kirschner, Sweller, and Clark (2006). *Educational Psychologist*, *42*(2), 91–97.

Sfard, A. (1998). On two metaphors for learning and the dangers of choosing just one. *Educational Researcher*, *27*(2), 4–13.

Sjøberg, S. (2018). The power and paradoxes of PISA: Should inquiry-based science education be sacrificed to climb on the rankings? *Nordic Studies in Science Education*, *14*(2), 186–202. doi:10.5617/nordina.6185

Stapleton, L. M., McNeish, D. M., & Yang, J. S. (2016). Multilevel and single-level models for measured and latent variables when data are clustered. *Educational Psychologist*, *51*(3–4), 317–330. doi:10.1080/00461520.2016.1207178

Windschitl, M., Thompson, J., & Braaten, M. (2008). Beyond the scientific method: Model-based inquiry as a new paradigm of preference for school science investigations. *Science Education*, 1–27. doi:10.1002/sce.20259

Wu, A. D., Li, Z., & Zumbo, B. D. (2007). Decoding the meaning of factorial invariance and updating the practice of multi-group confirmatory factor analysis: A demonstration with TIMSS data. *Practical Assessment, Research & Evaluation*, *12*(3), 1–26.

Zhang, L. (2016). Is inquiry-based science teaching worth the effort? *Science & Education*, *25*(7–8), 897–915. doi:10.1007/s11191-016-9856-0