

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

---

Papers in Natural Resources

Natural Resources, School of

---

2020

## Patterns of inquiry-based science instruction and student science achievement in PISA 2015

Cory Forbes

*University of Nebraska-Lincoln, cforbes3@unl.edu*

Knut Neumann

*IPN Leibniz Institute for Science and Mathematics Education, Kiel, Germany, neumann@leibniz-ipn.de*

Anja Schiepe-Tiska

*Technical University of Munich, schiepe-tiska@tum.de*

Follow this and additional works at: <https://digitalcommons.unl.edu/natrespapers>



Part of the [Natural Resources and Conservation Commons](#), [Natural Resources Management and Policy Commons](#), and the [Other Environmental Sciences Commons](#)

---

Forbes, Cory; Neumann, Knut; and Schiepe-Tiska, Anja, "Patterns of inquiry-based science instruction and student science achievement in PISA 2015" (2020). *Papers in Natural Resources*. 1385.

<https://digitalcommons.unl.edu/natrespapers/1385>

This Article is brought to you for free and open access by the Natural Resources, School of at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Papers in Natural Resources by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

# Patterns of inquiry-based science instruction and student science achievement in PISA 2015

Cory T. Forbes,<sup>1</sup> Knut Neumann,<sup>2</sup> and Anja Schiepe-Tiska<sup>3</sup>

1 University of Nebraska-Lincoln, Lincoln, NE, USA

2 IPN Leibniz Institute for Science and Mathematics Education, Kiel, Germany

3 Centre for International Student Assessment, Technical University of Munich, Munich, Germany

The authors share first authorship and appear in alphabetical order.

Correspondence — Cory T. Forbes [cforbes3@unl.edu](mailto:cforbes3@unl.edu)

## ORCID

Cory T. Forbes <http://orcid.org/0000-0003-2230-4251>

Knut Neumann <http://orcid.org/0000-0002-4391-7308>

## Abstract

Scientific literacy is a central aim of science education. The Programme for International Student Assessment (PISA) provides a measure of secondary students' scientific literacy and reported science instruction in 72 countries. Researchers have analyzed PISA data to identify important relationships between science instruction and students' science achievement. The purpose of this study is to further explore these relationships by using PISA 2015 data from a sample of 13 participating countries representing a range of mean science achievement. We use Latent Profile Analysis (LPA) to explore how students from highest- and lowest-performing profiles characterize the inquiry-based science instruction they experience. We then use cluster analysis

---

Published in *International Journal of Science Education* 2020, vol. 42, no. 5, pp. 783–806.

doi:10.1080/09500693.2020.1730017

Copyright © 2020 Informa UK Limited, trading as Taylor & Francis Group. Used by permission.

Submitted 25 May 2019; accepted 11 February 2020; published 18 Feb 2020.

Citation — Cory T. Forbes, Knut Neumann & Anja Schiepe-Tiska (2020) Patterns of inquiry-based science instruction and student science achievement in PISA 2015, *International Journal of Science Education*, 42:5, 783-806, DOI: 10.1080/09500693.2020.1730017

to compare patterns in inquiry-based instruction reported by students in the highest-performing profiles in each country and examine relationships between inquiry-based and other instructional practices. Findings from our analysis reveal fundamental differences in the frequency and nature of inquiry-based instruction reported by students associated with the highest and lowest levels of science achievement. For the highest performing profiles of students, results illustrate both consistencies and unique features of science instruction students report across four distinct clusters of countries. We discuss implications of these results for theory, research, and the design of science learning environments.

**Keywords:** PISA, secondary science, inquiry-based teaching, Elementary school, learning environment; models & modelling, elementary/primary school

## Introduction

In light of the urgency of today's most pressing global challenges, there has been increasing emphasis placed upon science education reform across K-12 contexts worldwide (e.g. Neumann, K., Fischer, H., & Kautertz, A., 2010; National Research Council [NRC], 2000, 1996; Waddington, Nentwig, & Schanze, 2007). Consistent with the global vision for science education, today's students must not only develop scientific understanding, but also learn to use it to reason, problem-solve, and make decisions about these challenges. Collectively, this knowledge and skills reflect an individual's *scientific literacy*. One instrument that provides a measure of scientific literacy is *The Programme for International Student Assessment* (PISA), a triennial international survey that assesses the knowledge and skills of 15-year-old students. PISA measures students' scientific literacy as the use of scientific knowledge to identify questions, acquire new knowledge, explain scientific phenomena, and draw evidence-based conclusions about science-related issues (OECD, 2017a, 2016a, 2016b).

But how can scientific literacy be best cultivated among students in science classrooms? Science education policy, both past and present, as well as international research, have and continue to theorize about and investigate the nature of science teaching and learning. One focus of this discussion has been on opportunities for students to engage in the processes and practices of science, or scientific inquiry. Historically contrasted with more traditional, teacher-centered instruction, a growing body of research indicates that effective inquiry learning is guided, with strategic direction from the teacher, as opposed to exclusively student-directed (e.g. Furtak, Seidel, Iverson, & Briggs, 2012; Minner, Levy,

& Century, 2010; Schroeder, Scott, Tolson, Huang, & Lee, 2007). Results from prior PISA analyses provide additional support for these findings (e.g. Hwang, Choi, Bae, & Shin, 2018; Lau & Lam, 2017; Tang, Tsai, Barrow, & Romine, 2019). Most studies based on PISA data, however, have focused on the relationship between science teaching and science achievement in a single country or compared science teaching between countries with different levels of student science achievement. Such research is limited by a lack of control of confounding variables at the country level (i.e. school system, educational expenditure, or cultural value of science education). More research is, therefore, needed on how groups of students with similar achievement levels *across* countries perceive science teaching and learning to support theoretical claims about the effect of inquiry-based teaching on students' scientific literacy.

To address these needs, we conducted analyses of data from the 2015 PISA administration. Specifically, we investigate relationships between reported science instruction and student outcomes in a subset of countries ( $n = 13$ ) from around of the world. These countries were purposefully selected to provide a representative sample of (a) levels of overall student science achievement and (b) contexts in which PISA items designed to measure inquiry-based teaching (the IBTEACH scale) do not exhibit a uni-dimensionality, suggesting comparative variation in how inquiry may be experienced by students in different countries. Using Latent Profile Analysis (LPA), we first identify groups of students reporting specific patterns of inquiry-based teaching and related these patterns to their science achievement in each of these countries. We then use cluster analysis to compare patterns of inquiry-based teaching associated with the highest average student achievement and investigate relationships between these patterns and other teaching variables (teacher-directed instruction, perceived feedback, and instructional adjustments) across countries.

## **Background and conceptual framework**

### ***Scientific literacy and PISA***

Science education reform policy has long advocated for and supported an emphasis on core knowledge of scientific disciplines, the processes and practices of science, and relationship between humans and the

natural world. While standards for science teaching and learning reflect these core areas of focus, collectively they speak to a more global goal of science education – scientific literacy – or abilities to apply scientific knowledge and practices to real-world phenomena in everyday life. PISA affords the opportunity to comparatively evaluate educational practices and outcomes, as well as related factors, through data collected from comparable samples of secondary students in over 70 countries. Science is a core discipline for which PISA provides a measure, focusing on core concepts of scientific domains (life, physical, Earth/space), how scientists engage in scientific study of the natural world to generate this knowledge, and the social, ontological, and cultural characteristics of these endeavors within and across scientific communities. The science competencies measured in PISA are collectively defined as ‘scientific literacy’, which reflects a skillset in response to the question, ‘What is important for young people to know, value and be able to do in situations involving science and technology?’ (OECD, 2017a, p. 20). More specifically, within PISA, scientific literacy is defined by three constituent dimensions: (1) explaining phenomena scientifically (content knowledge); (2) evaluating and designing scientific inquiry (procedural knowledge); and (3) interpreting data and evidence scientifically (epistemic knowledge) (OECD, 2017a), each of which align with core outcomes in standards for science teaching and learning worldwide. These domains are integral to the design of individual PISA science items.

### ***Science teaching and learning***

While the concepts inherent to scientific disciplines have always been the core focus of science education, in the last half-century, emphasis has grown on students’ participation in the processes and practices of science, or ‘science as inquiry’. While the particulars of these ideas are subject to ongoing theoretical and empirical discussion, the focus on ‘doing science’ has become conventional wisdom and an overarching, defining characteristic of the science education community, including practitioners, researchers, and policymakers, in the United States (NGSS Lead States, 2013; NRC, 1996) and internationally (e.g. Neumann, et al., 2010; Waddington et al., 2007). In PISA 2015, this emphasis is reflected in an inquiry-based teaching (IBTEACH) scale on the student questionnaire (OECD, 2016b, p. 242). Based on the meta-analyses of Furtak et al. (2012), inquiry-based science teaching is defined as a multifaceted

construct including different teaching activities such as conceptual, social, procedural, and epistemological activities. Some of these activities are more teacher-guided others more student-led (Müller, Prenzel, Seidel, Schiepe-Tiska, & Kjærnsli, 2016). These activities are reflected in nine ( $n = 9$ ) individual IBTEACH items reflecting two key dimensions of theoretical perspectives on inquiry. First, they foreground the kinds of inquiry practices in which students should engage, including identifying empirical questions, participating in experiments or investigations, working with data, formulating evidence-based claims about the natural world, and communicating and negotiating explanations for how and why natural phenomena occur. Second, they reflect the degree of student- or teacher-direction. Individual inquiry practices may be driven more heavily by students or be coupled with higher levels of teacher guidance. For example, students may design their own investigations (student-directed) or a teacher may provide students with a pre-determined investigation to conduct (teacher-directed). Both of these dimensions of classroom inquiry are critical components of the inquiry continuum (NRC, 2000) and relevant aspects of learning opportunities students are afforded in science classrooms.

Yet, long-standing questions persist about the evidence base upon which inquiry-based teaching rests. Science education researchers around the world have investigated process- and practice-based (i.e. inquiry-based) science teaching and learning spanning K-12 settings and disciplines, and this work has led to a significant literature base in the field of science education. However, scientific inquiry has historically been challenging to concretely define, operationalize, and measure through empirical research, and questions remain about which activities are most successful and about the degree to which these learning opportunities should be more student-led vs. teacher-guided. While the former has often been heralded as what Settlage (2007) describes as the 'gold standard' for science education, theoretical perspectives (Kirschner, Sweller, & Clark, 2006), as well as more recent meta-analyses and large-scale comparative studies on science education interventions (e.g. Furtak et al., 2012; Minner et al., 2010; Schroeder et al., 2007), suggest student-directed classroom inquiry may not be as effective as instruction involving significant guidance through science teachers. But how much and what types of guidance most strongly support students' science learning? These are important questions that merit further study, not only to define and describe the nature of instructional guidance that

may be most beneficial to students, but also to do so in varying and diverse secondary school contexts. PISA affords an opportunity investigate these questions using a largescale, comparative dataset from secondary science classrooms around the world that affords a measure of not only inquiry-based teaching, but also validated scales for teacher-directed teaching (TDTEACH), perceived feedback to students (PERFEED), and adjusting instruction in response to students' thinking (ADINST).

### ***Results of prior PISA-based research***

Many science education studies have been conducted using data from the 2006, 2009, and 2015 administrations of the PISA. While PISA data can only help establish associations between variables (not causal relationships), there are several consistent findings across these studies that are relevant to the study conducted here. First, while scientific inquiry is a core focus of science education reform, it is not the predominant or most common form of instruction students report experiencing in secondary science learning environments (OECD, 2016b; Tang et al., 2019). This is especially the case for its most student-directed derivation which, on average, is reported by less than 10% of students as the most common form of science instruction. For example, the majority (>70%) of students in OECD countries report being asked to design, conduct, and/or debate investigations and experiments only occasionally or never, while reported frequencies of sense-making activities, such as explanation, are far higher. In contrast, students report much higher incidence of teacher-directed science instruction, including teacher-led discussions, teacher explanations, and demonstrations. These findings suggest that most secondary science instruction worldwide, at least as students perceive it, is predominantly more teacher-directed with some integration of inquiry-based and more traditional pedagogical elements.

Second, prior research has shown that higher levels of reported student-directed inquiry is associated with lower overall scientific literacy. This is a universal finding in studies of the two most recent science-focused PISA administrations: 2006 and 2015. This trend has been well-documented based upon analyses of the 2006 data focused on the United States (Jiang & McComas, 2015), Australia, Canada, New Zealand (McConney, Oliver, Woods-McConney, Schibeci, & Maor, 2014), Finland (Kang & Keinonen, 2017), OECD countries in general (Kobarg et al., 2011), and many other countries (Cairns & Areepattamannil, 2017; Gee & Wong,

2012). Similar studies based upon the 2015 PISA data have investigated top performing countries, mostly in Asia and Europe (Aditomo & Klieme, 2020; Lau & Lam, 2017), lower-performing countries from South America and the Middle East (Hwang et al., 2018), the European Union (Costa & Araújo, 2018), and the United States (Tang et al., 2019). Across these studies, results show that the highest levels of secondary students' scientific literacy are associated with a combination of teacher- and student-directed instruction, often referred to as 'guided inquiry'. The assertion that scientific literacy is most strongly associated with instruction that includes both inquiry-based and more traditional instructional methods, but with significant structure and guidance from the teacher, is grounded in a substantial and growing body of evidence. With regard to different types of activities, particularly epistemic learning activities as well as a combination of procedural, epistemic, and social activities showed higher effect sizes as compared to traditional teaching that only emphasizes conceptual learning activities (Furtak et al., 2012).

However, despite this growing evidence base, a detailed account of inquiry-based instruction's underlying elements or consistency across contexts remains limited. Both student-led and teacher-guided inquiry instruction are constructs comprised of specific, underlying instructional practices represented by individual PISA questionnaire items which, in turn, define PISA's scales for inquiry-based teaching (IBTEACH). Despite overall reliability of these scales (OECD, 2017b), very recent studies based upon PISA 2015 suggest that IBTEACH may not always exhibit characteristics of a single scale (Aditomo & Klieme, 2020; Lau & Lam, 2017). Rather, the individual IBTEACH items may yield multiple sub-dimensions of inquiry-based instruction that can be interpreted as distinct from one another. For example, results of PISA 2006 analyses show that designing and conducting investigations may be a strong predictor of lower levels of scientific literacy, while sense-making about data and formulating explanations are more strongly associated with higher levels of scientific literacy (Areepattamannil, 2012; Areepattamannil, Freeman, & Klinger, 2011; Gee & Wong, 2012; Jiang & McComas, 2015; Kang & Keinonen, 2017), though each is part of the IBTEACH scale. However, while prior studies have focused on analysis of country-level data, there has been no prior research comparing these trends in subgroups of students at similar levels of science achievement across countries. This is an important area for future research to (a) establish which inquiry-based instructional practices are most strongly associated with science



achievement and (b) the degree to which these observed patterns are consistent for similar groups of students in different countries. To address this need, we ask the following research questions:

- (1) How can profiles of inquiry-based instruction associated with highest and lowest levels of students' science achievement be characterized?
- (2) For students in the profile associated with the highest levels of science achievement, how is the inquiry-based instruction they report similar or different between countries?
- (3) How are inquiry-based teaching patterns reported by students in the profile associated with the highest levels of science achievement related to other aspects of instruction?

## **Methods**

The main aim of this research is to further explore what types of inquiry activities are successful for students' learning and how much student-directed vs teacher-guided inquiry is necessary. Data used for this study were drawn from the 2015 administration of the PISA. Approximately 540,000 secondary students completed the assessment in 2015, representing about 29 million 15-year-olds in the schools of the 72 participating countries and economies. PISA provides a measure of science, reading, and mathematics through a 2-hour assessment and 35-minute questionnaire. On a rotating basis, one of these subject areas is a primary focus of PISA while the rest are minor domains. In 2015, science was the focus subject of PISA, meaning 1 hour of testing – 50% of the total assessment – was devoted to science, as well as a subset of items on a questionnaire related to science instruction.

## **Measures**

*Perceived science teaching.* As part of the student questionnaire, students were asked a series of questions about the science instruction they experience. We use four standardized indices of aspects of science instruction as characterized by students: Inquiry-based science teaching (IBTEACH), teacher-directed instruction (TDTEACH), perceived feedback from the teachers (PERFEED), and teachers' instructional adjustments (ADINST)

(see Appendix A; OECD, 2016b, p. 242). These indices have been the subject of previous research, were adapted to the specific needs of large-scale assessment studies (Müller et al., 2016), and exhibit high reliability ( $\alpha_{\text{TDTEACH}} = .82$ ,  $\alpha_{\text{IBTEACH}} = .86$ ,  $\alpha_{\text{PERFEED}} = .91$  and  $\alpha_{\text{ADINST}} = .80$ ).

*Science achievement/scientific literacy.* Items measuring science achievement in PISA 2015 (OECD, 2017b) were constructed around real-world scenarios, called ‘units’ designed around topics spanning different types of contexts (health, environment, natural resources, etc.) at varying levels of relevance to students (personal, community, global, etc.). Three item types are included for each unit: simple multiple choice, complex multiple choice, and constructed response items. The PISA 2015 science assessment contained a total 184 items. Collectively, these items comprise an overall scientific literacy score for students through a focus on context, knowledge, competencies, and attitudes related to science. Students spent one hour completing two, 30-minute ‘clusters’ of science items and no single student answered all items. Instead PISA employed a matrix-design, in which each student answers only a subset of the total number of items. As the focus of the study is not on individual student achievement but on the aggregated level of country, this design allows for a high level of accuracy by ensuring a short testing time for each student at the same time. That is, the PISA assessments solicits relatively few responses from each students on any one domain, while maintaining representation of a range of contexts, knowledge, competencies and attitudes. Hence, PISA employs a population model approach that is a combination of an Item Response Theory (IRT) model and a latent regression model. In PISA 2015, for each student, ten plausible values were estimated (see OECD, 2016a for scaling details). The PISA achievement scale has historically been set at 500 ( $SD = 100$ ). Since the PISA 2015 achievement scores are reported on a joint scale with previous PISA iterations to allow for trend analyses, the scale mean of the PISA 2015 iteration was 493 ( $SD = 94$ ).

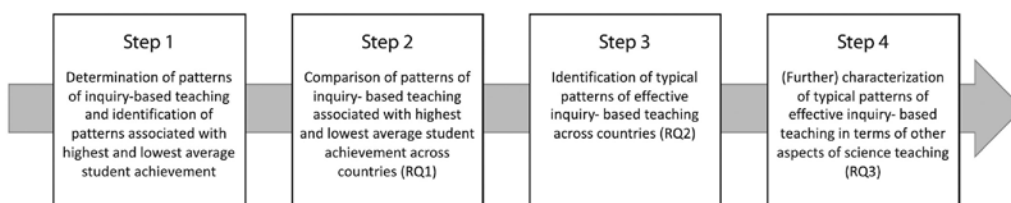
### ***Sample and data***

Here, we focus on analysis of student data ( $N = 74,877$ ) from 13 countries participating in PISA 2015. Selection criteria for these countries were twofold. First, these countries were selected to reflect a range of mean student science achievement levels below, at, and above the OECD

average. Second, they were selected because they exhibited variation within the IBTEACH scale. To establish the second selection criterion, we conducted exploratory factor analysis on data for each individual country. Results show that in each selected country, the IBTEACH items do not show evidence of a single scale but, rather, can be grouped into two or more distinct categories within each country (2 or more factors with Eigenvalues > 1). The exceptions to this selection criterion are [2 countries, The exceptions to this selection criterion are the United States and Germany, which were included in the analyses because they are the authors' home countries.], which were included in the analyses because [blinded for peer review]. For these countries, IBTEACH exhibits characteristics of single scale (a single factor with an Eigenvalue > 1), as it does for the majority of country-level PISA data. A summary of the PISA 2015 data from these countries used in the study is included in **Table 1**.

### *Data analysis procedures*

In order to answer our research questions, we determined typical patterns of inquiry-based teaching in each country and identified the patterns associated with the highest and lowest levels of student science achievement. We then compared patterns associated with highest and lowest science achievement across countries to explore if common characteristics of instructional patterns reported by high and low achieving students can be identified across countries (RQ1). Next, we examined instructional patterns reported by high achieving students in greater detail to identify typical patterns of effective inquiry-based teaching. Lastly, we explored how those patterns were related to other instructional variables such as teacher-directed teaching (TDTEACH), perceived feedback (PERFEED) and teachers' instructional adjustments (ADINST). **Figure 1** provides an overview of the steps performed in the data analysis.



**Figure 1.** Overview of data analysis steps.

**Table 1.** Descriptive statistics for country-level data (N = 13).

Country	N	Mean Science Achievement	Selection Tier	Gender		Grade <sup>a</sup>					
				Male	Female	7	8	9	10	11	12
Netherlands	5385	509	Above average	2685 49.9%	2700 50.1%	3 0.06%	152 2.82%	2242 41.63%	2944 54.67%	43 0.80%	1 0.02%
Japan	6647	538	Above average	3338 50.2%	3309 49.8%				6647 100%		
Korea	5581	516	Above average	2912 52.2%	2669 47.8%			523 9.37%	5032 90.16%	26 0.47%	
Denmark	7161	502	Above average	3559 49.7%	3602 50.3%	23 0.32%	1313 18.34%	5752 80.32%	73 1.02%		
Germany	6504	509	Above average	3307 50.8%	3197 49.2%	27 0.42%	486 7.47%	3063 47.09%	2827 43.47%	100 1.54%	1 0.02%
United States	5712	496	Average	2858 50.0%	2854 50.0%	1 0.02%	12 0.21%	529 9.26%	4210 73.70%	953 16.68%	7 0.12%
France	6108	495	Average	2997 49.1%	3111 50.9%	2 0.03%	54 0.88%	1353 22.15%	4487 73.46%	206 3.37%	6 0.10%
Czech Republic	6894	493	Average	3472 50.4%	3422 49.6%	28 0.41%	192 2.79%	2830 41.05%	3841 55.72%	3 0.04%	
Indonesia	6513	403	Below average	3170 48.7%	3343 51.3%	110 1.69%	425 6.53%	2581 39.63%	3245 49.82%	151 2.32%	1 0.02%
Mexico	7568	416	Below average	3765 49.7%	3803 50.3%	129 1.70%	281 3.71%	1916 25.32%	5184 68.50%	43 0.57%	15 0.20%
Hungary	5658	477	Below average	2808 49.6%	2850 50.4%	40 0.71%	209 3.69%	4542 80.28%	867 15.32%		
Bulgaria	5928	446	Below average	3105 52.4%	2823 47.6%	17 0.29%	120 2.02%	5536 93.39%	255 4.30%		
Romania	4876	435	Below average	2410 49.4%	2466 50.6%	45 0.92%	266 5.46%	3810 78.14%	755 15.48%		

a. Since PISA surveys a representative sample of 15 year-olds, depending on the country students from a range of different grades are sampled.

In the first step, to identify typical patterns of inquiry-based teaching, we employed Latent Profile Analysis. Latent Profile Analysis (LPA) is a statistical method used to identify a series of homogeneous subgroups in a sample with respect to a series of continuous indicators (Oberski, 2016). In order to determine the number of subgroups for each country with respect to the IBTEACH items, we followed the commonly used procedure to fit a series of models assuming one, two, three, etc. up to ten subgroups and identifying the best fitting model (Holzberger, Praetorius, Seidel, & Kunter, 2019). In Appendix B, we present the Likelihood (LL), Akaike Information Criterion (AIC), Bayes Information Criterion (BIC), and Entropy as well as the Lo-Mendell-Rubin (LMR), Vuong-Lo-Mendell-Rubin (VLMR) and Bootstrapped Likelihood Ratio Tests (BLRT). We identified the optimal number of profiles for each country based upon minimum LL, AIC and BIC, significance in LMR, VMLR and BLRT, and most importantly, a local maximum in Entropy – ideally above .90 or at least .80, which are commonly considered as indicating very good or good fit (Celeux & Soromenho, 1996). Following this procedure, we obtained the following number of profiles for the respective countries (marked bold in Table S2): Netherlands (4), Japan (4), Korea (3), Denmark (5), Germany (5), USA (3), France (5), Czech Republic (4), Indonesia (5), Mexico (3), Hungary (3), Bulgaria (4), Romania (4). We continued our analysis by determining the subgroups with the highest and lowest levels of student science achievement and respective pattern of inquiry instruction (i.e. the average score on each IBTEACH item). **Table 2** shows the descriptive statistics for the highest and lowest achieving subgroups for each of the countries, which range in size from a few hundred to over 3000 students.

In the second step, for an initial examination of the inquiry teaching patterns reported by high and low achieving students, we compared average scores across items for each pattern by means of t-tests. To investigate whether patterns differ across countries, we used Multivariate Analyses of Variance (MANOVA) followed by subsequent t-tests; that is, we determined whether there were any differences between patterns associated with the highest and the lowest achievement, before we identified the specific countries that differed and on which items these countries differed. A power analysis suggests that at an average sample size of  $N = 1812$  the MANOVA can detect already extremely small effects of less than .01, and ANOVA or t-tests respectively can detect very small

**Table 2.** Descriptive statistics for highest- and lowest-performing students by country (N = 13).

<i>Country</i>	<i>Profile</i>	<i>N</i>	<i>Mean</i>	<i>Standard deviation</i>
Netherlands	High	923	551.59	85.86
	Low	477	456.68	97.79
Japan	High	1308	565.25	87.62
	Low	559	498.58	101.01
Korea	High	3119	539.16	88.9
	Low	494	464.91	95.92
Denmark	High	950	527.02	81.19
	Low	220	458.61	82.67
Germany	High	1354	547.99	94.66
	Low	601	486.92	94.93
United States	High	2120	506.99	97.31
	Low	630	464.73	92.6
France	High	570	545.17	85.25
	Low	400	463.42	93.38
Czech Republic	High	1020	508.8	90.16
	Low	712	457.46	94.09
Indonesia	High	2846	412.96	67.29
	Low	384	368.1	66.4
Mexico	High	3244	422.06	70.49
	Low	1163	399.72	70.68
Hungary	High	2227	490.69	88.71
	Low	451	425.81	94.67
Bulgaria	High	2315	481.87	94.71
	Low	462	399.35	91.28
Romania	High	1564	443.29	76.67
	Low	287	404.21	82.43

effects of .12. To minimize the risk of a Type I error and misidentification of non-significant differences as statistically-significant, we have employed Bonferroni correction were appropriate.

In the third step, to identify typical patterns of effective inquiry-based teaching, we utilized cluster analysis to group country-level subgroups showing similar patterns of inquiry-based instruction associated with the highest level of student achievement. We determined the number of significant clusters using hierarchical clustering with the assumption of no a priori groups. Hierarchical clustering is a method in which clusters are build bottom-up; that is, initially, each pattern is considered to form a unique cluster, then the two clusters that are considered most similar (or least different for that matter) are grouped into a single cluster. This latter step is repeated until all patterns form a single cluster.

Hierarchical clustering requires determining (dis-) similarities (i.e. distances) between clusters. In case of a metric scale, which is the best approximation to the averaged values of student ratings on the items, the most commonly used distance measure is Euclidean. Hence, we chose this measure as the distance measure in our analysis. In addition to a distance measure, a clustering (i.e. aggregation method) is needed. Here, we use the *complete linkage* method (Whitaker & Christman, 2014), which aims to find the largest distance between the elements of the two clusters, since we were aiming to identify similar profiles. Since meaningful clustering of profiles across thirteen countries naturally leads to small sample sizes (i.e. smaller number of profiles grouped into each cluster), we relied exclusively on descriptive statistics in comparing clusters of profiles instead of applying inferential statistics.

In the fourth and last step, in order to further characterize patterns of effective inquiry-based instruction, we examined the average scores of teacher-directed teaching (TDTEACH), students' perceived feedback (PERFEED) and teachers' instructional adjustments (ADINTS) for the clusters identified.

The data were analyzed using the statistical software *R* 3.5.2 (R Core Team, 2018a) and *MPLUS* (Muthén & Muthén, 2018). The publicly-available SPSS data were loaded into R using the package *foreign* (R Core Team, 2018b). We used the *MplusAutomation* (Hallquist & Wiley, 2018) package to run the Latent Profile Analysis. In order to determine average student achievement for the subgroups identified based on plausible values accounting for student weighing we utilized the *BIFIEsurvey* package (Robitzsch & Oberwimmer, 2018). For the cluster analysis we utilized the *clustsig* R-package (Whitaker & Christman, 2014).

## Results

In the results that follow, we present findings for each of our stated research questions. First, we identify and describe patterns of inquiry-based instruction as characterized by students associated with the highest and lowest levels of student science achievement across countries. Next, we examine differences in patterns of inquiry-based instruction between countries for the patterns associated with the highest levels of student science achievement. Finally, we characterize broader instructional

patterns within previously identified groups of countries sharing similar patterns using additional PISA indices for science instruction (TDTEACH, PERFEED, ADINST).

### ***Inquiry-based instructional profiles for highest- and lowest-performing students***

In research question #1, we asked, 'How can profiles of inquiry-based instruction associated with highest and lowest levels of students' science achievement be characterized?' To address research question #1, we conducted LPA to identify the inquiry-based teaching pattern associated with the highest and lowest levels of student science achievement for each country. Findings from these analyses are presented in **Figure 2**, which shows the average scores on the items that comprise the IBTEACH scale for each of these two patterns across countries. In Figure 2, the lines represent the mean while the shaded area reflects the range across all thirteen countries. These results show that in inquiry-based instruction associated with the lowest student achievement, every single aspect that constitutes inquiry-based teaching occurs in most to all lessons. Accordingly, when accounting for alpha level error inflation, only the difference between the frequency of *students explaining their ideas* (Q01) and *students spending time doing experiments* (Q02) differ significantly from one another,  $t(20.01) = -4.42, p < .0014$ . No statistically-significant differences are observed in pairwise comparisons of the remaining IBTEACH items for the lowest-achieving subgroup. In other words, the profile associated with the lowest mean student science achievement across these 13 countries is characterized by high levels of all aspects of inquiry-based teaching.

In contrast, the pattern of inquiry-based instruction associated with highest levels of student science achievement differs from the pattern associated with lowest levels of student science achievement in two ways. Students reporting this pattern report comparatively less frequent experiences with all aspects of inquiry, as evidenced by the lower mean item scores for all IBTEACH items. Results of a multivariate analysis of variances (MANOVA) test indicate that the differences in patterns of inquiry-based teaching reported by high and low achieving students are statistically-significant,  $F(1,9) = 47.87, p < .001$ . Pairwise comparisons of individual IBTEACH item means in the two patterns show that all





**Figure 2.** Patterns of inquiry-based teaching for high and low achieving students across countries based on students' ratings of the frequency of occurrence (1-never or hardly ever, 2-in some lessons, 3-in most lessons, 4-in all lessons) of IBTEACH items (ST098) Q1 through Q10: Student explain their ideas (Q01), Students spend time doing experiments (Q02), Students argue about science questions (Q03), Students draw conclusions from experiments (Q05), Teacher explains ideas (Q06), Students design their own experiments (Q07), Class debate about investigations (Q08), Teacher explains relevance (Q09), Students test ideas (Q10).

observed differences for all items are statistically significant (Appendix C). In the majority of cases the difference ranges between all vs. most (e.g. students explaining their ideas, Q01) or some to many lessons (e.g. *students spending time doing experiments*, Q02). The largest difference can be observed for *students designing their own experiments* (Q07) which is substantially lower for the pattern associated with high achievement as compared to the pattern associated with low achievement. Additionally, much more internal variation is observed in IBTEACH items within the pattern associated with highest levels of student science

**Table 3.** Significance of T-tests between IBTEACH items (ST098) means within profile associated with highest levels of student science achievement.

ST098	Q01	Q02	Q03	Q05	Q06	Q07	Q08	Q09	Q10
Student explain their ideas	Q01	*				***	***		*
Students spend time doing experiments	Q02				*				
Students argue about science questions	Q03								
Students draw conclusions from experiments	Q05					**			
Teacher explains ideas	Q06					***	***		**
Students design their own experiments	Q07							***	
Class debate about investigations	Q08							*	
Teacher explains relevance	Q09								
Students test ideas	Q10								

\*  $p < .0015$ \*\*  $p < 0.00028$ \*\*\*  $p < 0.000028$ 

Significance levels have been adjusted using Bonferroni correction for multiple T-tests.

achievement as compared to the pattern associated with lowest levels of student science achievement. Even when correcting for alpha level error inflation, statistically significant differences are observed between mean scores for several IBTEACH items WITHIN the pattern associated with highest levels of student science achievement, as shown in **Table 3**. For example, the lowest means are observed for more student-directed and investigation-focused items, including *students designing their own experiments* (Q07) and *class debate about investigations* (Q08). Differences are statistically significant between these lower item means and more teacher-guided, conceptual, and sense-making-focused items, including *students explain their ideas* (Q01), *teacher explains ideas* (Q06), and *teacher explains relevance* (Q09), which have the highest items means within the profile associated with highest levels of student science achievement. Overall, these findings would suggest that patterns of inquiry-based teaching associated with the highest levels of students' science achievement are characterized by more differentiated instruction in which different aspects of inquiry are highlighted or emphasized comparatively more or less across periods of instruction.

In summary, these results illustrate overarching patterns in inquiry-based teaching (IBTEACH) as characterized by students in these 13 countries who, on average, exhibited the highest- and lowest levels of science achievement. Students reporting the pattern associated with

lowest levels of student science achievement note more frequent and consistent experiences with all aspects of inquiry-based instruction. In contrast, students reporting the pattern associated with highest levels of student science achievement indicate relatively less frequent experiences with inquiry-based instruction, particularly for student-led procedural activities, such as *designing their own investigations* (Q07) and *testing their own ideas* (Q10), as well as *arguing about science questions* (Q03). Explanation- focused, conceptual and social activities are most frequently reported while investigation- focused, procedural practices are relatively less commonly reported in both patterns.

### ***Inquiry-based instruction across countries for highest-performing students***

In research question #2, we asked, 'For students in the profile associated with the highest levels of science achievement, how is the inquiry-based instruction they report similar or different between countries?' To address research question #2, we disaggregate the data for students in the pattern associated with highest levels of student science achievement (Figure 2) by country. More specifically, we investigated the individual patterns associated with the highest average student achievement for each country and compared the profiles across countries. Mean scores for IBTEACH items are presented for each country in **Table 4**. This information provides an overview of the average occurrence of each aspect of inquiry-based teaching as reported by students associated with the highest achieving pattern for each of these thirteen countries. These results corroborate the aggregate findings from research question #1 for the pattern associated with the highest levels of student science achievement. Not only is variation observed between IBTEACH items within data for individual countries but, as importantly, these patterns vary between countries. What this suggests is that the inquiry-based instruction experienced by students reporting the pattern associated with the highest achievement in each country differ from one another. Students report, for example, that they are allowed to *explain their ideas* (Q01) in most to all lesson in Romania, whereas students in Korea report them being allowed to explain their ideas in only some lessons,  $t(3711.11) = 56.347, p < .001$ . Students in France *design their own experiments* in only few lessons, but *spend time doing experiments* in many lessons,

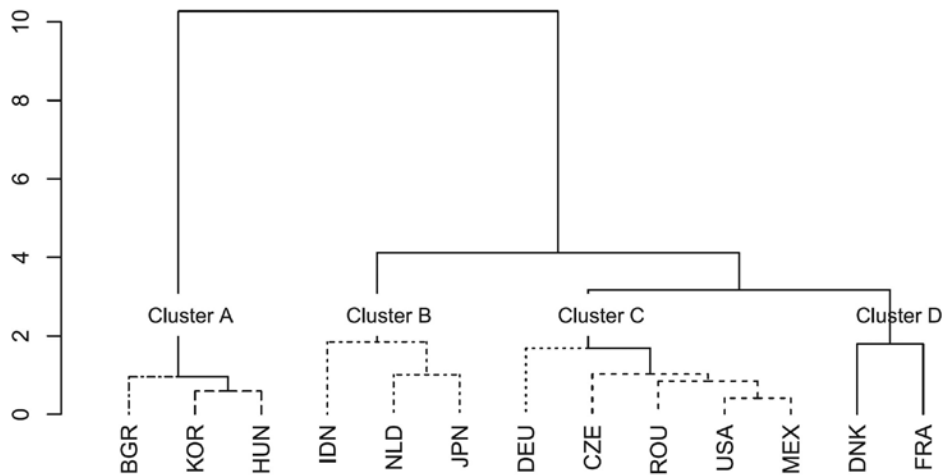
**Table 4.** IBTEACH item means and standard deviations by country for highest-performing profile.

Item	Profile	Country <sup>a</sup>												
		NLD	JPN	KOR	DNK	DEU	USA	FRA	CZE	IDN	MEX	HUN	BGR	ROU
Q01 Student explain their ideas	High	2.89	2.82	2.04	3.52	3.31	3.35	3.62	3.52	2.94	3.41	2.35	2.36	3.47
	Low	0.74	0.95	0.92	0.65	0.69	0.65	0.63	0.71	0.87	0.70	0.92	0.85	0.76
Q02 Students spend time doing experiments	High	3.23	3.33	3.42	3.87	3.37	3.92	3.84	3.69	3.70	3.88	3.57	3.93	3.80
	Low	0.73	0.68	0.55	0.36	0.74	0.33	0.44	0.56	0.60	0.36	0.63	0.35	0.59
Q03 Students argue about science questions	High	2.51	2.18	1.22	3.35	2.08	2.55	2.70	2.13	1.73	2.32	1.18	1.26	2.21
	Low	0.80	0.89	0.44	0.61	0.68	0.75	0.73	0.78	0.56	0.80	0.41	0.48	0.76
Q05 Students draw conclusions from experiments	High	2.99	2.62	3.02	3.83	2.74	3.56	3.05	2.51	2.80	3.48	2.80	3.39	3.28
	Low	0.86	1.05	0.85	0.44	0.87	0.74	1.00	1.04	1.01	0.79	0.99	0.97	0.94
Q06 Teacher explains ideas	High	2.02	1.65	1.12	3.39	2.36	2.34	3.51	2.71	2.67	2.26	1.38	1.84	2.44
	Low	0.77	0.83	0.35	0.71	0.76	0.84	0.58	0.84	0.84	0.75	0.56	0.64	0.82
Q07 Students design their own experiments	High	2.79	2.86	3.24	3.84	2.89	3.58	3.64	3.07	3.58	3.61	3.03	3.82	3.57
	Low	0.94	0.96	0.64	0.43	0.89	0.78	0.64	0.90	0.73	0.62	0.91	0.43	0.68
Q08 Class debate about investigations	High	2.80	2.84	1.18	3.63	3.08	3.07	3.76	3.16	2.23	2.80	1.41	1.58	2.78
	Low	0.70	0.95	0.41	0.56	0.74	0.68	0.45	0.84	0.63	0.81	0.59	0.61	0.86
Q09 Teacher explains relevance	High	3.06	3.20	3.24	3.88	3.18	3.88	3.82	3.10	3.65	3.74	3.13	3.80	3.64
	Low	0.77	0.82	0.65	0.39	0.71	0.37	0.42	0.88	0.62	0.51	0.85	0.47	0.65
Q10 Students test ideas	High	2.93	3.06	2.44	3.78	3.00	3.29	3.74	3.48	2.94	3.34	2.31	2.26	3.06
	Low	0.67	0.74	0.98	0.44	0.76	0.66	0.50	0.64	0.86	0.72	0.90	0.79	0.87
Q08 Class debate about investigations	High	3.03	3.30	3.42	3.90	3.10	3.92	3.85	3.39	3.64	3.83	3.34	3.90	3.62
	Low	0.78	0.68	0.53	0.30	0.74	0.33	0.41	0.72	0.64	0.45	0.78	0.34	0.74
Q09 Teacher explains relevance	High	1.00	1.47	1.11	1.48	1.00	2.11	1.75	1.47	1.89	2.06	1.07	1.27	1.63
	Low	0.00	0.74	0.36	0.50	0.00	0.84	0.75	0.52	0.75	0.88	0.27	0.46	0.48
Q10 Students test ideas	High	3.29	2.76	3.21	3.80	3.29	3.44	3.57	3.39	3.40	3.31	3.21	3.80	3.62
	Low	0.46	1.04	0.69	0.40	0.46	0.92	0.71	0.54	0.84	0.93	0.68	0.51	0.60
Q08 Class debate about investigations	High	1.66	1.00	1.06	3.00	2.53	2.25	2.01	2.38	1.44	2.29	1.31	1.62	2.30
	Low	0.67	0.00	0.25	0.83	0.76	0.81	0.86	0.82	0.50	0.81	0.52	0.64	0.86
Q09 Teacher explains relevance	High	2.91	3.34	3.34	3.84	3.12	3.75	3.65	3.27	3.60	3.55	3.35	3.89	3.53
	Low	0.87	0.47	0.54	0.42	0.75	0.54	0.64	0.74	0.51	0.71	0.71	0.33	0.72
Q10 Students test ideas	High	2.79	2.73	2.14	3.55	2.50	3.07	2.65	3.25	3.39	3.23	2.15	2.15	3.13
	Low	0.76	0.93	0.93	0.60	0.87	0.74	0.93	0.72	0.70	0.75	0.93	0.80	0.81
Q10 Students test ideas	High	2.98	3.26	3.36	3.81	2.95	3.88	3.69	3.39	3.67	3.78	3.38	3.87	3.68
	Low	0.82	0.72	0.57	0.47	0.84	0.35	0.61	0.68	0.60	0.50	0.69	0.38	0.56
Q10 Students test ideas	High	1.82	1.54	1.12	2.75	2.25	2.77	1.81	2.96	2.09	2.81	1.36	1.42	2.44
	Low	0.80	0.78	0.34	0.92	0.77	0.75	0.78	0.74	0.81	0.82	0.60	0.57	0.90
Q10 Students test ideas	High	2.86	3.01	3.29	3.72	2.98	3.88	3.59	3.15	3.49	3.66	3.10	3.79	3.56
	Low	0.87	0.89	0.62	0.57	0.81	0.38	0.67	0.82	0.72	0.62	0.86	0.53	0.68

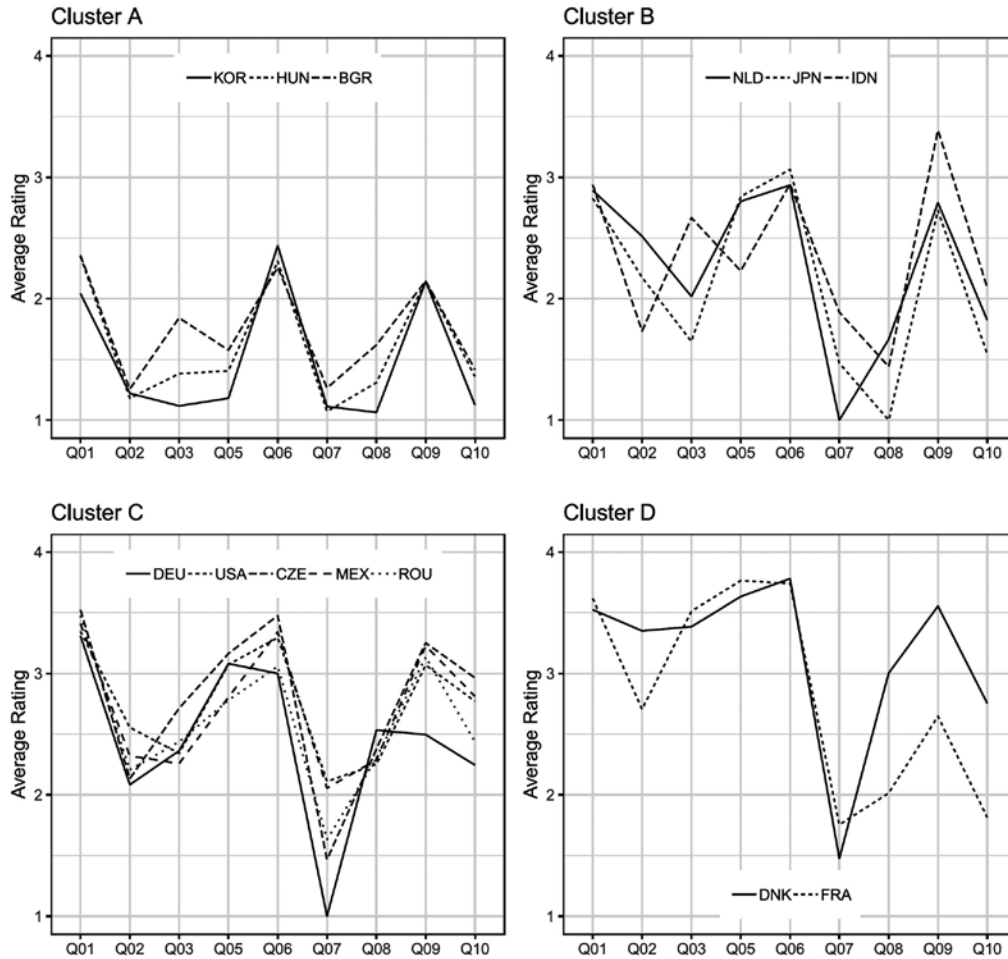
a. NLD, Netherlands; JPN, Japan; KOR, Korea; DNK, Denmark; DEU, Germany; USA, United States; FRA, France; CZE, Czech Republic; IDN, Indonesia; MEX, Mexico; HUN, Hungary; BGR, Bulgaria; ROU, Romania.

$t(1123.50) = 21.57, p < .001$ , whereas students in Bulgaria both *spend time doing experiments* and *design their experiments* in only few lessons,  $t(4493) = 0.66, p = .51$ . These findings suggest that the inquiry-based teaching associated with the highest levels of student science achievement differ across countries. Results of a MANOVA of the nine IBTEACH items across countries show that there are statistically-significant differences between countries in terms of IBTEACH item means,  $F(12, 108) = 419, p < .001$ , which is expected given the large sample sizes of the country-specific data ( $570 < n < 3244$ , see data analysis section).

In order to further explore similarities or dissimilarities between IBTEACH item means between countries, and to attempt to more formally establish patterns of effective inquiry-based instruction across countries, we performed a cluster analysis based on comparisons of the country-specific IBTEACH item data. **Figure 3** shows the resulting dendrogram, which displays the similarity or (increasing) dissimilarity respectively between data for individual countries. The ideal number of clusters is commonly determined by identifying larger gaps between levels of dissimilarity (at the point(s) where patterns are merged into a cluster). In the dendrogram shown in Figure 3, this applies to the height of approximately 2.0, suggesting the country-specific IBTEACH data are best grouped into four different clusters. The patterns for these clusters are shown in **Figure 4**.



**Figure 3.** Dendrogram resulting from cluster analysis grouping profiles by similarity of IBTEACH across countries: Netherlands (NLD), Japan (JPN), Korea (KOR), Denmark (DNK), Germany (DEU), United States (USA), France (FRA), Czech Republic (CZE), Indonesia (IND), Mexico (MEX), Hungary (HUN), Bulgaria (BGR), Romania (ROU).



**Figure 4.** Country-specific IBTEACH groupings identified from cluster analysis of IBTEACH items (ST098) Q1 through Q10: Student explain their ideas (Q01), Students spend time doing experiments (Q02), Students argue about science questions (Q03), Students draw conclusions from experiments (Q05), Teacher explains ideas (Q06), Students design their own experiments (Q07), Class debate about investigations (Q08), Teacher explains relevance (Q09), Students test ideas (Q10); and countries: Netherlands (NLD), Japan (JPN), Korea (KOR), Denmark (DNK), Germany (DEU), United States (USA), France (FRA), Czech Republic (CZE), Indonesia (IND), Mexico (MEX), Hungary (HUN), Bulgaria (BGR), Romania (ROU).

Figure 4 shows four distinct clusters of countries representing patterns of inquiry-based instruction associated with highest average levels of student science achievement. Each pattern is defined by the relatively infrequent occurrence of *students design their own investigations* (Q07), which was shown previously in response to RQ#1. However,

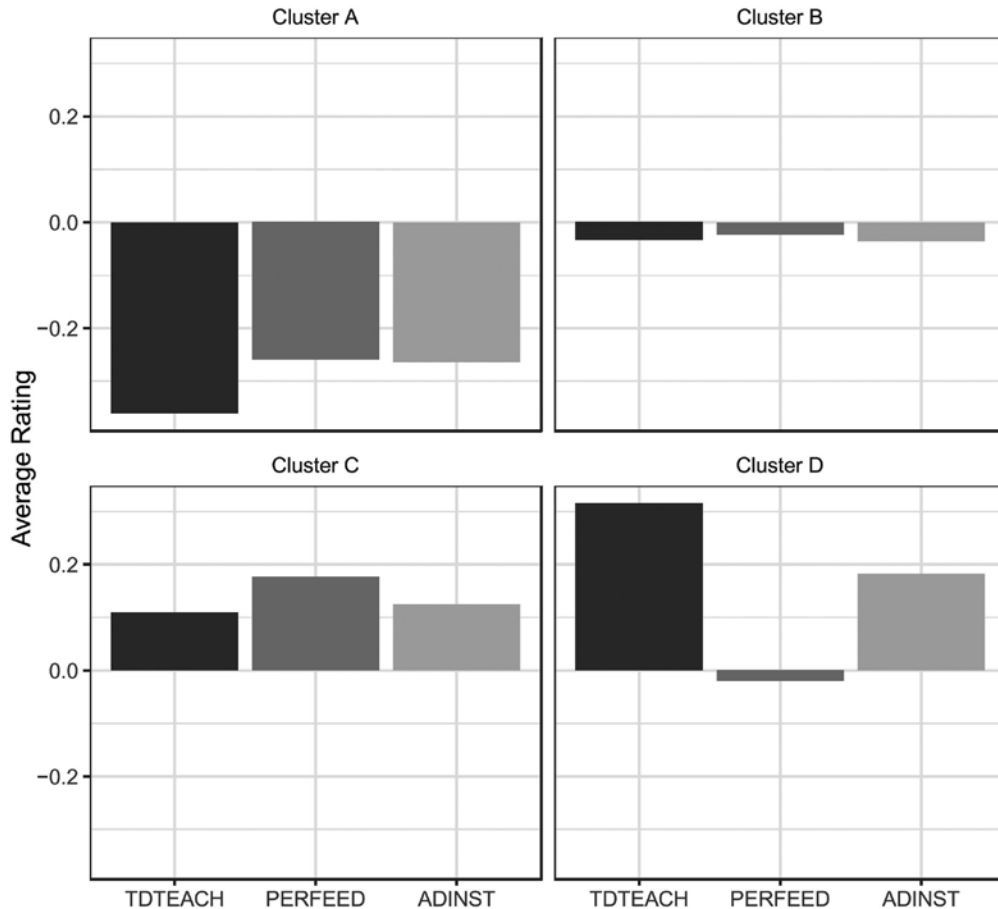
these patterns are distinguishable from one another in a number of different ways. Cluster A is arguably the most distinct, characterized by overall low occurrences of all aspects of inquiry-based teaching as compared to the other profiles. However, within Cluster A, students consistently report relatively more frequent occurrences of conceptual learning activities such as *teachers explaining ideas* (Q06), and *teachers explaining the relevance* of the respective ideas (Q09) as well as more frequent occurrence of *students explaining their ideas* (Q01). Never or hardly ever do they report engaging in other social, procedural, or epistemic learning activities. Comprised of students from Bulgaria, Korea, and Romania, it includes countries that, overall, have mean science achievement levels above and below the OECD average. In contrast to Cluster A, Cluster D is characterized by high frequencies of almost every aspect of inquiry-based instruction, though mean scores are comparatively lower for *students testing their ideas* (Q10) and *classroom debate about investigations* (Q08). While sharing an emphasis on explanation, inquiry-based instruction in the countries in Cluster D also focus on *students argue about science questions* (Q03), the epistemic activity of *students drawing conclusions from experiments* (Q05) and *spend time in the laboratory doing practical experiments* (Q2). This smallest cluster, Cluster D, is comprised of Denmark and France, which fall above and at the OECD average, respectively. Clusters B and C are less distinct from one another. Both clusters are characterized by moderate (to high) levels of conceptual and social activities, both by students and the teacher. Scores for conceptual and social explanation-focused IBTEACH items (Q01, 06, 09) in Cluster C are nearly as high as observed in Cluster B. The most notable difference between Clusters B and C, however, is that students in Cluster B report little to no *classroom debate about investigations* (Q08) and somewhat lower levels of *students testing their ideas* (Q10). Both Clusters B and C may be interpreted to represent guided inquiry, with moderate levels of student involvement and teacher guidance but subtle differences in the role students testing their ideas through investigations and the discussion of those investigations play. Cluster C includes the highest number of countries of any of the clusters and both Clusters B and C each include countries that fall above, at, and below the OECD mean for science achievement.

In summary, results of analyses in response to RQ#2 show observed differences in patterns of inquiry-based instruction associated with the highest levels of student science achievement. While the relative de-emphasis on *students design their own investigations* (Q07) was consistent across all countries, observed clusters of countries exhibited similarities and differences from one another based upon the IBTEACH items. Four distinct clusters were identified, each of which largely differed from the others based upon the (a) overall frequency of reported inquiry practices and (b) comparative frequencies of IBTEACH items within the cluster. Individual clusters were comprised of countries in which overall science achievement scores were above, at, and below the OECD average, suggesting these observed patterns of inquiry-based instruction are not necessarily associated with overall science achievement levels in a particular country. Overall, these findings illustrate shared characteristics and unique patterns of inquiry-based instruction as reported by students in profiles associated with the highest levels of science achievement in these countries. They suggest that inquiry-based instruction associated with the strongest evidence of students' science learning may vary by country.

### ***Correlational patterns with instructional variables across countries***

In research question #3, we asked, 'How are inquiry-based teaching patterns reported by students in the profile associated with the highest levels of science achievement related to other aspects of instruction?' To address research question #3, we examined relationships between IBTEACH index values for clusters of countries identified in RQ#2 and other PISA instructional variables, including teacher-directed instruction (TDTEACH), perceived feedback (PERFEED), and teachers' instructional adjustments (ADINST). To report results of this analyses, we present mean values for these indices for each country cluster identified in response to RQ#2 as deviations from the mean values for of TDTEACH ( $M= 2.55, SD = 0.29$ ), PERFEED ( $M= 2.10, SD = 0.28$ ), and ADINST ( $M = 2.49, SD = .29$ ) across all 13 countries for students in the profile associated with the highest levels of science achievement. Results of these analyses are presented in **Figure 5**.





**Figure 5.** Cluster-specific deviations from group mean values for teacher-directed teaching (TDTEACH), perceived feedback to students (PERFEED), and adjusting instruction in response to students' thinking (ADINST) scales.

Overall, observed mean values for TDTEACH, IBTEACH, and ADINST associated with these four clusters of countries follow a consistent pattern. Reported frequencies of these three other instructional practices generally mirror those reported by these same students for IBTEACH in terms of frequency. Clusters C and D exhibit reported higher frequencies for TDTEACH, PERFEED, and ADINST than the overall sample mean, while those in Cluster A and B report frequencies for these forms of instruction that are at or below the sample mean. Reported frequencies of TDTEACH and ADINST are highest for Cluster D, which is aligned with its overall high reported frequencies of inquiry-based teaching also reported by students in Cluster D. Cluster A, in which students generally

reported the least frequent exposure to inquiry-based instruction, also exhibits the lowest relative levels of TDTEACH, PERFEED, and IDINST. Clusters B and C are similarly aligned with relatively moderate frequencies of teacher-directed teaching, perceived feedback, and adaptive instruction consistent with frequencies of IBTEACH reported by these same students. However, in Cluster C, students report the highest level of perceived feedback (PERFEED). Also, students in countries represented by Clusters A, B, and C also report lower or similar frequencies of teacher-directed instruction (TDTEACH) as compared to both perceived feedback and adjusting instruction.

However, one cluster – Cluster D – represents a partial departure from this overall trend. Students in countries represented by Cluster D (Denmark and France) report higher levels of teacher-directed instruction than PERFEED or ADINST. Also, while comparatively high observed levels of TDTEACH and ADINST are directly aligned with frequencies of IBTEACH reported by students in these four clusters, reported frequencies of perceived feedback for students in Cluster D are lower than reported by students in Cluster C and fall below the overall mean value for PERFEED amongst students in the highest performing profiles across all 13 countries. Unlike the other three clusters, students in Denmark and France report far less perceived feedback than teacher-directed instruction as compared to the overall sample mean. Despite exhibiting the most frequent occurrence of teacher-directed teaching and adaptive instruction, students in Cluster D report some of least frequent exposure to perceived feedback from their teachers.

In summary, results of analyses in response to RQ#3 suggest that overall reported levels of instructional practices for science generally mirrored those students reported for inquiry-based teaching. The more frequent inquiry-based instruction students reported experiencing, the more frequent they also reported their science teachers engaging in teacher-directed instruction, perceived feedback, and adaptive instruction. High-performing students in Denmark and France, however, reported (a) comparatively higher levels of teacher-directed teaching than students in the other countries and (b) less frequent experience with perceived feedback than would have been expected given overall trends in reported science instruction across all four clusters of students from profiles associated with the highest levels of science achievement.

## Discussion

Findings from this study illustrate observed trends in inquiry-based teaching as reported by groups of students associated with the highest and lowest relative levels of science achievement across countries in which the index for inquiry-based teaching (IBTEACH) does not operate as a single scale as measured in PISA 2015. First, there is an ongoing need to better understand science instruction as perceived by students exhibiting varying levels of science achievement. Study findings show that inquiry-based instruction reported by these groups of secondary students differs substantially. In response to RQ#1, we found that the pattern associated with highest levels of student science achievement, as opposed to the pattern associated with lowest levels of student science achievement, (a) is characterized by less frequent experiences with inquiry-based instruction and (b) exhibits more internal variation in specific inquiry practices. These findings generally reinforce findings from previous PISA studies (Areepattamannil, 2012; Areepattamannil et al., 2011; Cairns & Areepattamannil, 2017; Gee & Wong, 2012; Jiang & McComas, 2015; Kang & Keinonen, 2017), including recent studies focused on PISA 2015 (Aditomo & Klieme, 2020; Hwang et al., 2018; Lau & Lam, 2017; Tang et al., 2019) and meta-analyses of previous inquiry-based interventions (e.g. Furtak et al., 2012; Minner et al., 2010; Schroeder et al., 2007), which have shown that high levels of inquiry-based teaching are relatively uncommon and associated with lower student science achievement while higher levels of student science achievement are associated with relatively more teacher-directed forms of inquiry. They have also provided evidence for relative frequencies of specific inquiry-based instructional practices, showing that highly student-driven dimensions of inquiry, particularly procedural activities associated with investigation, are least frequently associated with high levels of student science achievement while the most frequently reported were explanation-focused conceptual and epistemic inquiry practices focused on sense-making. Results presented here extend prior research by providing evidence that this trend is consistent across countries, independent of overall observed mean levels of science achievement at the country level, even when the precise nature of inquiry-based instruction (IBTEACH) varies between

countries. These empirical results lend strength to the argument for the importance of teacher guidance in science learning environments and provide insight into which of these instructional practices may be most beneficial for students' science learning.

Second, given the observed association between teacher guidance and highest levels of student science achievement, it is also important to understand how guided inquiry can vary for groups of high-achieving students. Study findings show that while overall patterns of inquiry-based teaching illustrate consistencies discussed previously, at the country level, inquiry-based instruction can also differ, as evidenced by the four clusters identified through LPA in response to RQ#2. In these four clusters, inquiry-based instruction varies in terms of (a) reported frequencies of specific inquiry practices and (b) distinct configurations in which specific inquiry practices are more or less prominent, largely independent of other factors, such as country-level mean science achievement. For example, France and Denmark (Cluster D), with average and above-average mean science achievement, not only exhibit the highest overall levels of reported inquiry-based teaching, but disproportionately so for specific social and epistemic inquiry practices of arguing about questions and drawing conclusions from evidence. In contrast, Korea, Bulgaria, and Hungary (Cluster A), with above and below average mean levels of science achievement, illustrate the lowest levels of reported inquiry practices and, specifically for the items referenced for Cluster D, by a factor of two. Though PISA's IBTEACH index is intended to provide a single, concrete measure of inquiry-based instruction (OECD, 2017b), results presented here show that it does not always exhibit uni-dimensional behavior for subsets of data. This finding has also been observed in other recent studies from PISA 2015 focused on highest and lowest-performing countries (Aditomo & Klieme, 2020; Lau & Lam, 2017). However, study findings contribute to this work by showing that this variation is not simply a function of overall science achievement levels. Rather, we may hypothesize that these observed variations in inquiry-based teaching for highest-performing groups of students may be the result of an array of factors, such as country-level science curricula, cultural norms of teaching and learning, or the disciplinary course contexts of PISA participants. Further research could explore differences between

these clusters utilizing other PISA variables, such as courses of student respondents and items from the teacher questionnaire. While it is not possible to provide causal explanation for *why* these variations in inquiry-based teaching exist using PISA data, they do suggest that the nature of what may be tentatively identified as *effective* inquiry-based instruction can vary by context and region.

Finally, third, better understanding secondary science teachers' domain-general instructional practices may help provide insight into these patterns of inquiry-based instruction and more precisely account for the type of guidance teachers provide. Results show that the inquiry-based instruction students report experiencing exhibit relationships to other forms of science instruction, such as teacher direction, feedback, and adaptive instruction, that are both consistent and unique. As shown in response to research question #3, the more frequent experiences with inquiry-based science instruction students report, the higher the frequencies of teacher-directed instruction, perceived feedback, and adaptive instruction they also report. However, countries in one cluster (D; Denmark and France), illustrate patterns of instruction that defy this trend, with lower than expected frequencies of perceived feedback. Few previous PISA studies have investigated interrelationships between these instructional indices from the PISA student questionnaire in regard to science. However, those that have suggest that teacher-directed and adaptive instruction are generally associated with higher levels of student science achievement, while perceived feedback is negatively associated with students' science achievement (Costa & Araújo, 2018; Lau & Lam, 2017). Results presented here generally reinforce these findings for teacher-directed and adaptive instruction, but do not necessarily indicate the same negative correlation with perceived feedback. Our findings indicate that for the highest-performing group of students across these 13 countries, perceived feedback may be as effective as other instructional practices when used in parallel with inquiry-based instruction. Furthermore, the observable and significant variation observed in perceived feedback for Denmark and France suggests that this instructional practice, while generally productive for the highest-achieving groups of students, may also be highly variant across all groups of students in all countries.

## Limitations and conclusion

We recognize specific limitations of the study. While PISA affords a very large and representative dataset of science education worldwide, it is nonetheless limited in other ways. First, the data for instructional practices (IBTEACH, TDTEACH, PERFEED, and ADINST) is entirely self-reported by students. The quality of this data can be influenced by many factors, including students' interpretation of the individual items and their individual motivation to complete the assessment accurately. Second, it is not possible to establish causal relationships between variables of interest. In this study, we are unable to claim that observed patterns of inquiry-based instruction directly impact or result in observed levels of students' science achievement. Rather, PISA data is descriptive, affording the opportunity to characterize the nature of associations between variables, but not direct relationships. Furthermore, results of this study do not afford insight into other factors that may be associated with either student science achievement or instruction. Previous studies have explored other student-level variables, suggesting students' motivation, interest in science, and socio-economic profile are also important factors in science instruction they report and their scientific literacy. Future work should investigate the characteristics of students in the profiles associated with highest and lowest levels of student science achievement, whether they are also similar across countries, the degree to which they mediate observed relationships between reported science instruction and student science achievement presented in this study. Additionally, differences between numbers of students, mean science achievement scores, and standard deviations of the highest- and lowest-performing student subgroups at the country level suggest that the groups of students reporting these teaching practices are not identical, though they are comparable. More research is, therefore, needed to understand how and why science teaching practices, particularly forms associated with the highest levels of student science achievement, may vary in particular countries, regions, and/or contexts.

Despite these limitations, study results build upon previous, PISA-based science education research (Areepattamannil, 2012; Areepattamannil et al., 2011; Cairns & Areepattamannil, 2017; Gee & Wong, 2012; Jiang & McComas, 2015; Kang & Keinonen, 2017; Kobarg et al., 2011;

McConney et al., 2014), including recent studies focused on PISA 2015 (Aditomo & Klieme, 2020; Hwang et al., 2018; Lau & Lam, 2017; Tang et al., 2019), and yield important findings that contributes to the field's broader understanding of the nature of secondary science instruction and its relationship to students' scientific literacy. Though reported associations are correlative, they show that the science instruction characterized by the highest performing groups of students in individual countries exhibit marked similarities, even when overall mean levels of science achievement for these countries are above, at, and below the OECD average. In essence, effective science instruction – science instruction associated with the highest levels of students' science achievement – exhibits both shared characteristics across contexts and comparative variation that allows for localized implementation. The use of LPA and cluster analysis to examine similar groups of students across countries that reflect a range of science achievement but each share variation within the IBTEACH scale is a novel approach.

Scientific literacy has been, and will continue to be, a foundational goal of international science education efforts (e.g. Neumann, et al., 2010; NRC, 2000, 1996; Waddington et al., 2007). However, the more challenging question is how to optimally foster this outcome in formal classroom settings through science instruction. PISA data, with its large and representative international sample, provides one important tool to address this question. Study findings specifically add evidence to the growing body of literature that has illustrated the benefits of guided inquiry over both traditional, direct instruction and entirely student-directed inquiry learning (e.g. Furtak et al., 2012; Minner et al., 2010; Schroeder et al., 2007) and contribute to theoretical perspectives on the nature of science instruction (Kirschner et al., 2006; Settlage, 2007). These empirical insights highlighting the kinds of inquiry practices associated with highest levels of student science achievement can provide important guidance for science teachers, science curriculum developers, teacher educators, and professional developers working to help shape secondary science learning environments. They also reinforce the importance of context specific science teaching and learning, where secondary science learning environments can be cultivated to exhibit these general characteristics of inquiry-based teaching but also to afford flexibility for local norms, whether cultural, curricular, and/or procedural in nature, to shape effective science instruction. Ultimately, in these ways,

the translation of science education research into practice is an important role that PISA data can serve in pursuit of international science education efforts and the global scientific literacy.



**Disclosures** No potential conflict of interest was reported by the authors.

**Funding** This work was supported by the Fulbright Scholar Program, a program of the United States Department of State, Bureau of Educational and Cultural Affairs. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the U.S. government. The authors share first authorship and appear in alphabetical order. We thank the teachers and students who made this research possible.

## References

- Aditomo, A., & Klieme, E. (2020). Forms of inquiry-based science instruction and their relations with learning outcomes: Evidence from high and low-performing education systems. *International Journal of Science Education*. <https://doi.org/10.1080/09500693.2020.1716093>
- Areepattamannil, S. (2012). Effects of inquiry-based science instruction on science achievement and interest in science: Evidence from Qatar. *Journal of Educational Research*, *105*(2), 134–146.
- Areepattamannil, S., Freeman, J., & Klinger, D. (2011). Influence of motivation, self-beliefs, and instructional practices on science achievement of adolescents in Canada. *Social Psychology of Education*, *14*(2), 233–259.
- Cairns, D., & Areepattamannil, S. (2017). Exploring the relations of inquiry-based teaching to science achievement and dispositions in 54 countries. *Research in Science Education*, *49*, 1–23.
- Celeux, G., & Soromenho, G. (1996). An entropy criterion for assessing the number of clusters in a mixture model. *Journal of Classification*, *13*, 195–212.
- Costa, P., & Araújo, L. (2018). Quality of teaching and learning in science (EUR 28865 EN). Publications office of the European Union. *Luxembourg*. doi:10.2760/860512.JRC109064.
- 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.



- Gee, K. A., & Wong, K. K. (2012). A cross national examination of inquiry and its relationship to student performance in science: Evidence from the Program for International Student Assessment (PISA) 2006. *International Journal of Educational Research*, 53, 303–318.
- Hallquist, M. N., & Wiley, J. F. (2018). Mplusautomation: An R package for facilitating large-scale latent variable analyses in Mplus. *Structural Equation Modeling*, 1–18.
- Holzberger, D., Praetorius, A.-K., Seidel, T., & Kunter, M. (2019). Identifying effective teachers: The relation between teaching profiles and students' development in achievement and enjoyment. *European Journal of Psychology of Education*, 34, 801–823.
- Hwang, J., Choi, K. M., Bae, Y., & Shin, D. H. (2018). Do teachers' instructional practices moderate equity in mathematical and scientific literacy? An investigation of the PISA 2012 and 2015. *International Journal of Science and Mathematics Education*, 16(1), 25–45.
- Jiang, F., & McComas, W. F. (2015). The effects of inquiry teaching on student science achievement and attitudes: Evidence from propensity score analysis of PISA data. *International Journal of Science Education*, 37(3), 554–576.
- Kang, J., & Keinonen, T. (2017). The effect of student-centered approaches on students' interest and achievement in science: Relevant topic-based, open and guided inquiry-based, and discussion-based approaches. *Research in Science Education*, 48, 865–885.
- 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.
- Kobarg, M., Prenzel, M., Seidel, T., Walker, M., McCrae, B., Cresswell, J., & Wittwer, J. (2011). *An international comparison of science teaching and learning: Further results from PISA 2006*. Münster: Waxmann.
- Lau, K.-C., & Lam, T. Y.-P. (2017). Instructional practices and science performance of 10 top-performing regions in PISA 2015. *International Journal of Science Education*, 39(15), 2128–2149.
- McConney, A., Oliver, M. C., Woods-McConney, A., Schibeci, R., & Maor, D. (2014). Inquiry, engagement, and literacy in science: A retrospective, cross-national analysis using PISA 2006. *Science Education*, 98(6), 963–980.
- 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 years 1984 to 2002. *Journal of Research in Science Teaching*, 47(4), 474–496.
- 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). New York, NY: Springer.
- Muthén, L. K., & Muthén, B. O. (2019). *Mplus user's guide* (8th ed.). Los Angeles, CA: Muthén & Muthén.

- National Research Council. (1996). *National science education standards*. Washington, DC: National Academy Press.
- National Research Council. (2000). *Inquiry and the national science education standards: A guide for teaching and learning*. Washington, DC: The National Academies Press.
- Neumann, K., Fischer, H., & Kauertz, A. (2010). From PISA to educational standards: The impact of large-scale assessments on science education in Germany. *International Journal of Science and Mathematics Education*, 8(3), 545–563.
- NGSS Lead States. (2013). *Next generation science standards. For states, by states*. Washington, DC: The National Academies Press.
- Oberski, D. (2016). Mixture models: Latent profile and latent class analysis. In *Modern statistical methods for HCI* (pp. 275–287). Cham: Springer.
- OECD. (2016a). *PISA 2015 results (volume I): Excellence and equity in education*. Paris: OECD Publishing.
- OECD. (2016b). *PISA 2015 results (volume II): Policies and practices for successful schools*. Paris: OECD Publishing.
- OECD. (2017a). *PISA 2015 assessment and analytical framework: Science, reading, mathematics, financial literacy and collaborative problem solving*. Paris: OECD Publishing.
- OECD. (2017b). *PISA 2015 technical report*. Paris: OECD Publishing.
- R Core Team. (2018a). Foreign: Read data stored by 'Minitab', 'S', 'SAS', 'SPSS', 'Stata', 'Systat', 'Weka', 'dBase'. Retrieved from <https://CRAN.R-project.org/package=foreign>
- R Core Team. (2018b). *R: A language and environment for statistical computing*. Retrieved from <https://www.R-project.org/>
- Robitzsch, A., & Oberwimmer, K. (2018). *BIFIEsurvey: Tools for survey statistics in educational assessment*. Retrieved from <https://CRAN.R-project.org/package=BIFIEsurvey>
- Schroeder, C. M., Scott, T. P., Tolson, H., Huang, T. Y., & Lee, Y. H. (2007). A meta-analysis of national research: Effects of teaching strategies on student achievement in science in the United States. *Journal of Research in Science Teaching*, 44(10), 1436–1460.
- Settlage, J. (2007). Demythologizing science teacher education: Conquering the false ideal of open inquiry. *Journal of Science Teacher Education*, 18(4), 461–467.
- Tang, N.-E., Tsai, C.-L., Barrow, L., & Romine, W. (2019). Impacts of enquiry-based science teaching on achievement gap between high-and-low SES students: Findings from PISA 2015. *International Journal of Science Education*, 41(4), 448–470.
- Waddington, D., Nentwig, P., & Schanze, S. (2007). *Making it comparable. Standards in science education 1664*. Münster: Waxmann.
- Whitaker, D., & Christman, M. (2014). *Clustsig: Significant cluster analysis*. R Package version 1.1.2014. Retrieved from <https://cran.hafro.is/web/packages/clustsig/clustsig.pdf>