

This is a self-archived version of an original article. This version may differ from the original in pagination and typographic details.

Author(s): Saarela, Mirka; Kärkkäinen, Tommi

Title: Knowledge Discovery from the Programme for International Student Assessment

Year: 2017

Version:

Copyright: © Springer International Publishing AG 2017

Rights: In Copyright

Rights url: <http://rightsstatements.org/page/InC/1.0/?language=en>

Please cite the original version:

Saarela, M., & Kärkkäinen, T. (2017). Knowledge Discovery from the Programme for International Student Assessment. In A. Peña-Ayala (Ed.), *Learning Analytics : Fundamentals, Applications, and Trends. A View of the Current State of the Art to Enhance e-Learning* (pp. 229-267). Springer International Publishing. *Studies in Systems, Decision and Control*, 94. https://doi.org/10.1007/978-3-319-52977-6_8

Chapter 8

Knowledge Discovery from the Programme for International Student Assessment

Mirka Saarela¹, Tommi Kärkkäinen²

¹ University of Jyväskylä, Department of Mathematical Information Technology
40014 Jyväskylä, Finland
mirka.saarela@jyu.fi

² University of Jyväskylä, Department of Mathematical Information Technology
40014 Jyväskylä, Finland
tommi.karkkainen@jyu.fi

Abstract. The Programme for International Student Assessment (PISA) is a worldwide study that assesses the proficiencies of 15-year-old students in reading, mathematics, and science every three years. Despite the high quality and open availability of the PISA data sets, which call for big data learning analytics, academic research using this rich and carefully collected data is surprisingly sparse. Our research contributes to reducing this deficit by discovering novel knowledge from the PISA through the development and use of appropriate methods. Since Finland has been the country of most international interest in the PISA assessment, a relevant review of the Finnish educational system is provided. This chapter also gives a background on learning analytics and presents findings from a novel case study. Similar to the existing literature on learning analytics, the empirical part is based on a student model; however, unlike in the previous literature, our model represents a profile of a national student population. We compare Finland to other countries by hierarchically clustering these student profiles from all the countries that participated in the latest assessment and validating the results through statistical testing. Finally, an evaluation and interpretation of the variables that explain the differences between the students in Finland and those of the remaining PISA countries is presented. Based on our analysis, we conclude that, in global terms, learning time and good student-teacher relations are not as important as collaborative skills and humility to explain students' success in the PISA test.

Keywords: PISA, Learning Analytics, Big Data, Knowledge Discovery, Hierarchical Clustering

Abbreviations

ESCS Economic, social, and cultural status
LA Learning Analytics
MOOC Massive open online course
OECD Organisation for Economic Cooperation and Development
PISA Programme for International Student Assessment

8.1 Introduction

The original purpose of *Learning Analytics* (LA), as stated by researchers such as Siemens (2013, p. 1383) and Ferguson (2012, p. 306), was to “measure, collect, analyze, and report data about learners and their contexts, for the purposes of understanding and optimizing learning and the environments in which it occurs.” Slightly different variants were later offered to characterize the discipline (Pardo & Teasley 2014, Gray et al. 2014, Siemens & Baker 2012). Increased attention to Massive Open Online Courses (MOOCs) (e.g., Wang et al. 2014, Ye & Biswas 2014, Reich et al. 2014, Coffrin et al. 2014, Hickey et al. 2014, Santos et al. 2014, Vogelsang and Ruppertz 2015, Ferguson and Clow 2015, Hansen and Reich 2015, Wise et al. 2016, Hecking et al. 2016) has intensified the need for data-based learning support from the perspective of *big data*. This is evidenced by several articles (e.g., Picciano 2012, Chatti et al. 2012, Siemens 2012, Chatti et al. 2014, Dawson et al. 2014, Wise & Shaffer 2015, Merceron et al. 2016) as well as by the theme of the 2015 Learning Analytics and Knowledge conference “Scaling Up: Big Data to Big Impact” (see Dawson et al. 2015).

The PISA is a worldwide triennial survey conducted by the Organisation for Economic Cooperation and Development (OECD), resulting in publicly available educational data on a large scale. In addition to assessing the proficiency of 15-year-old students from different countries and economies in reading, mathematics, and science, the PISA provides “data about learners and their contexts” as one of the largest public databases¹ of students’ demographic and contextual data, such as their attitudes and behaviors toward various aspects of education. More than seventy countries and economies have already participated in the PISA, and the assessment is referred to as the “world’s premier yardstick for evaluating the quality, equity, and efficiency of school systems” (OECD 2013a).

In the PISA studies, data collection is of very high quality, including the development of the appropriate instruments, the procedures, and the storage of the data in public databases. This is evidenced by the large amount of money spent on ensuring quality related to these issues. However, much less money has been invested in the analysis of the collected data, and only a few PISA analysis studies have resulted in publications in the scientific field (Olsen 2005a). Rutkowski et al. (2010) argue that the size of the PISA data sets as well as the technical complexities within them may be the reason why more researchers do not work with these freely available and high-quality data.

Our research is motivated by the lack of secondary analysis of the PISA data, which calls for the development and utilization of big data LA methods for making discoveries within the international domain of the PISA. Such methods can then be used to summarize the PISA data sets in novel ways in order to better understand students from diverse countries and the settings in which they learn (Siemens & Baker 2012). Hence, in relation to big data LA, we focus on the inter-

¹The PISA data can be downloaded from <http://www.oecd.org/pisa/pisaproducts/>.

national context in an effort to understand national education systems as learning environments. Such a scope for LA was also emphasized by Long & Siemens (2011), who pointed out that LA should occur on the national and international levels, primarily targeting national governments and education authorities. As a classroom is in a school is in a city is in a region is in a country is in a continent, thorough use of educational data and empirical evidence should be linked to those principles and practices of educational systems that are known to have an effect on learning. This is the primary concern in the PISA.

Chatti et al. (2014) introduced a reference model for LA based on four dimensions (stakeholders, objectives, data, and methods) that resembles the critical LA dimensions suggested by Greller & Drachsler (2012). Fig. 8.1 illustrates how large-scale educational assessments, such as the PISA, can leverage big data LA according to these dimensions. Specifically, national bodies introduce the objectives (i.e., the factors that constitute good national education systems) for assessing the international student population. Then, large amounts of data representing student background and proficiency are sampled and transformed into derived representations, whose characteristics (the sample to population alignment introducing weights and the rotated test design introducing missing values) must be handled by applied LA methods. When meaningful patterns are found, these are reported back to the educational decision makers.

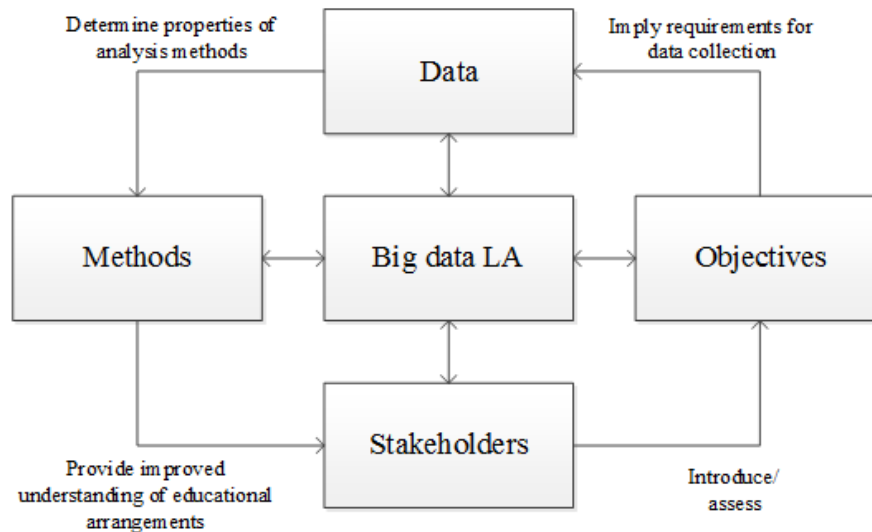


Fig. 8.1 Conducting big data LA for large-scale educational system assessments (cf. Chatti et al. 2014, Greller & Drachsler 2012).

Ferguson et al. (2014) emphasize the large-scale institutional adoption of appropriate educational patterns. In the best case, the institutional meso-level approaches are aggregated from the upscale local micro-level patterns and from the downscale macro-level characteristics of a good educational system. Thus, mean-

ingful patterns at the macro level (e.g., within a large educational organization) originate from characteristics of a large student population in relation to the rigorously measured learning outcomes.

The structure of this paper is as follows. In Section 8.2, we provide necessary background on big data LA and educational knowledge discovery from the PISA. In Section 8.3, a relevant review on methodologically related studies is provided, and the forms and complexities of PISA data are described. Next, the overall analysis method is depicted in Section 8.4. In Section 8.5, the results and interpretations of the hierarchical clustering of the aggregated country profiles are presented and statistically validated. In Section 8.6, the PISA results are visualized in a dashboard. Finally, in Section 8.7, the empirical work is summarized, and in Section 8.8, the overall conclusions are presented.

8.2 Background and Related Work

We next provide the necessary theoretical background for the empirical part of the chapter. First, we explain big data LA and summarize LA methods. Then, we characterize a pool of methodologically related work on the use of clustering in educational data analysis. We observe that methodologically related studies are typically conducted on the micro level of individual courses or tutoring systems.

8.2.1 Toward Big Data LA

As emphasized in the introduction, LA studies are increasingly leveraging big data. The term “big” in “big data” does not solely refer to the amount of data but actually references four “V”s (the first three according to Laney (2001) and the last one as described by Gupta et al. (2014)): (i) *Volume* refers to the size of data sets caused by the number of data points, their dimensionality, or both; (ii) *Velocity* is linked to the speed of data accumulation; (iii) *Variety* stands for heterogeneous data formats, which are caused by distributed data sources, highly variable data gathering, etc.; and (iv) *Veracity* refers to the fact that (secondary) data quality can vary significantly, and manual curation is typically impossible.

In relation to big data LA, PISA data are characterized by a high volume and low veracity due to missing values, but there is no velocity and small, well-managed variety due to the meticulous design. Moreover, unlike the existing LA studies, the collected student sample is aligned to the whole worldwide student population under study using weights (see the last paragraphs in Section 8.2.3). For example, the sample data of the PISA 2012 consists of approximately half a million students, representing 24 million 15-year-old students from 68 different countries and territories.

Chatti et al. (2012) state that different LA techniques for detecting interesting educational patterns originate from four analysis categories: *statistics*; *information visualization*; *data mining* (identifying this with knowledge discovery in databases) in the form of classification, clustering, and association rule mining; and *social network analysis*. Other LA researchers support this notion that data mining and knowledge discovery techniques are one category of the broader set of LA methods. Rogers (2015), for example, lists data mining as one of the more sophisticated quantitative methods in LA, and Siemens (2013) states that knowledge discovery from databases is an LA technique that has become increasingly important.

Generally, with the advent of big data in education, LA methods have shifted from the more traditional data analysis techniques, such as statistics, to more scalable data mining methods (Hershkovitz et al., 2016; Joksimović et al., 2016). In fact, Ferguson (2012) points out that the two main differences between general educational research and the specific research field of LA (according to the LA definition given in the beginning of this chapter) is that LA “make[s] use of pre-existing, machine-readable data, and that its techniques can be used to handle ‘big data.’”

Application of data mining and knowledge discovery methods in an educational context typically realizes an *educational knowledge discovery process* that, especially when using an open educational data set like that of the PISA, supports learning and knowledge analytics (Verbert et al. 2012). Several case studies (e.g., Hu et al., 2016; Brown et al., 2016; Grawemeyer et al., 2016; Allen et al., 2016; Chandra & Nandhini 2010) have proven the need for and the success of specific knowledge discovery processes and data analysis methods within the educational domain. However, data from many of the existing educational case studies are specific to certain educational environments or institutions, which complicates the comparison of the techniques and the results provided.

In contrast, the PISA tests are standardized, and the resulting data sets are comparable between different nations and their educational arrangements. Hence, the PISA provides an interesting and novel case for big data LA techniques (Saarela & Kärkkäinen 2014, Saarela & Kärkkäinen 2015a,b,c, Kärkkäinen & Saarela 2015), combining the methodological requirements that are due to the above-mentioned technical complexities of the data with comparative educational knowledge discovery.

8.2.2 On Educational Data Analysis Using Clustering

As has been pointed out above, *clustering* is one of the key techniques in the data mining category of the LA methods. Next, we describe a pool of work related to the clustering of educational data as well as the empirical work in Sections 8.4–8.5. This set of papers was primarily identified by scanning through the most relevant publication forums (see Saarela et al. 2016a) in the field, especially the *Jour-*

*nal of Learning Analytics*² and the *Conference on Learning Analytics & Knowledge*³, restricting the topic to clustering with real educational data sets. The description of the work is organized according to the clustering method used and the size of the clustered educational data set.

Hierarchical Clustering. Logs of 454 online mathematics practice sessions by 69 students were clustered by Desmarais & Lemieux (2013). In that study, preprocessing first transformed the logs into temporal sequences (time series) reflecting the state of interaction between the student and the learning environment. These representations were then clustered using an agglomerative hierarchical method, and the interpretation of the result was based on visualizing the clusters as state sequence diagrams. Three characteristic forms of using the system were identified: (i) exploratory browsing, (ii) short practice sessions, and (iii) exercise-intensive sessions.

Self-regulatory strategies of undergraduate students, especially their characteristics in accessing online learning material, were studied by Colthorpe et al. (2015). Hierarchical clustering of 97 students was able to separate high- and low-performing students, and the low-performing students were characterized by extensive use of lecturing recordings. This could, however, be explained by the form of engagement with the learning material.

Segedy et al. (2015) provided a more in-depth analysis of students' self-regulated interaction with the learning material in an open-ended computer-based learning environment. Student assessment was based on the coherence analysis, whose descriptive metrics for 99 sixth grade students were separated into five clusters using complete-link hierarchical clustering as part of the versatile analysis process. In addition to two very small clusters of (i) confused guessers and (ii) students disengaged from the task, the main clusters characterized the self-regulated interaction patterns of (iii) frequent researchers and careful editors, (iv) strategic experimenters, and (v) engaged and efficient students.

Hu et al. (2016) used hierarchical clustering to analyze the responses of 523 English and Chinese primary school students to a questionnaire about their reading behaviors, reading preferences, and attitudes toward reading. Three main reading profiles were identified, and they were fully characterized by good, moderate, and bad reading habits.

Hecking et al. (2016) combined social similarity (i.e., distances in the communication graph of the students) and semantic similarity (i.e., distances between the content-based roles by the students) to construct a socio-semantic block modeling approach for analyzing a MOOC discussion forum. Hierarchical clustering was used in the actual construction of the block model from the derived similarity measure. The analysis of the communication graph of 647 students in 502 threads on 27 forums verified the presence of different roles, with a moderate correlation

² See <http://learning-analytics.info/>.

³ See <http://lakXX.solaresearch.org/>, where XX stands for year in which the conference took place. For example, <http://lak16.solaresearch.org/> contains a link to the proceedings of the 2016 conference.

between the social and the semantic role by a student. Discovery of the three main socio-semantic roles suggested that online discussion forums need better recognition and adaptation to the different user roles.

K-Means. A collaboration of 31 participants in a math discussion board was addressed by Xing et al. (2014) through the lens of activity theory, which links individual and social behavior, using the prototype-based k-means clustering method. In this study, the important phases of the educational clustering process, preprocessing, and interpretation of the clustering result were strongly present. The result consisted of three clusters characterizing (i) learners who were personally participative but less communicative on the group level, (ii) collaboratively participating but shallow learners, and (iii) less participative poor learners.

An automated approach using the k-means clustering algorithm was described by Li et al. (2013) for constructing a student model from the content features of algebra problems. Methodologically versatile preprocessing (feature extraction, min-max scaling, and principal component analysis) and tenfold cross-validation characterized the approach. The experiment with data from 71 students concluded that the clustering-based model was at least as good as the prior manually constructed model, as it was able to reveal previously unidentified and valuable knowledge components of mathematical problem solving. An innovative assessment of the physical learning environment that also used the k-means clustering method was reported by Almeda et al. (2014). The result consisted of four different clusters characterizing the similar content profiles of 30 classroom walls, as decorated by the teachers.

Multiple clustering methods (including k-means and hierarchical clustering) at various stages of the data analysis were applied by Blikstein et al. (2014) to reveal the different patterns and trends of the development of programming behavior in an introductory undergraduate programming course. The overall analysis of 370 participants and 154,000 code snapshots was concluded in multiple ways. First, for different tasks within LA, different kinds of tools are needed, ranging from fast and simple wrap-ups of data to advanced machine-learning methods running on high-performance computing platforms. Secondly, concerning the clustering methods, it is necessary to have either better support to interpret the result of a clustering method or the application of more advanced methods to improve the potential insights and knowledge discovery from data. Thirdly, concerning the domain of the study, the changes in the code update patterns by the students were more strongly correlated with the course performance compared to the size of code updates.

A subset of methods used by Blikstein et al. (2014) were also utilized by Worsley & Blikstein (2014) to analyze the problem-solving patterns of 13 students for open-ended engineering tasks. This LA method was based on the segmentation and extraction of action features from the hand-coded video data. The k-means algorithm produced four clusters whose interpretation could be summarized into two principal dimensions of idea quality and design process, which were both related to students' level of experience.

Expectation Maximization. Bouchet et al. (2013) clustered the derived variables of multiple thematic groups from the log data of 106 college students using an intelligent tutoring system fostering self-regulated learning. They used the expectation-maximization algorithm from Weka, resulting in three clusters as suggested by the knee point (see Saarela & Kärkkäinen 2015a), after careful cross-validation with multiple restarts. The three clusters were generally characterized by varying levels of performance but also reflected (through metadata) differences in the number of self-regulated learning processes in which the students were engaged. Bogarin et al. (2014) also used the expectation-maximization algorithm from Weka and discovered three clusters from the log data of 84 Psychology students training to learn online with Moodle. In particular, a cluster of the most passive online students was detected, of which two-thirds failed the course.

Activity in online discussion forums as a predictor of study success was also studied by López et al. (2012). Methodologically, it was shown that the prototypes obtained from the expectation-maximization clustering algorithm with tenfold cross-validation with Weka software were able to distinguish 114 different and informative cases of university student behavior. Similar to Bogarin et al. (2014), it was concluded that active participation in the course forum was a good predictor of the final grade for the course.

Summary. To summarize this small survey of educational clustering methods, hierarchical clustering, k-means, and expectation maximization were the most common approaches. This was also the conclusion in the review by Peña-Ayala (2014). Similarly, student modeling, including behavior and performance models, was the dominant educational data analysis approach, covering all the assessed research except Almeda et al. (2014) (see Table 11 in the work published by Peña-Ayala 2014). Note that a set of older references concerning the use of clustering in educational settings, as briefly introduced by Bouchet et al. (2013) in Section 6, also emphasized the student model as an important part of intelligent online tutoring systems.

8.2.3 *LA Approaches Oriented to Analyze PISA Repositories*

As concluded in the previous section, clustering is one of the key techniques for analyzing educational data, especially in LA. However, most of the educational clustering studies use small data sets of tens or at most hundreds of students at the micro and meso levels of educational systems. By comparison, the PISA 2012 data set is comprised of around half a million students and represents a population of 24 million people worldwide (see the last paragraph in Section 8.3.2).

A considerable amount of literature has been published on the PISA. However, as observed by Olsen (2005a), these publications are mainly national or international reports that have not undergone the peer-review process. Furthermore, many of the peer-reviewed publications dealing with the PISA (e.g., Deng & Gopinathan 2016, Auld & Morris 2016, Rasmussen & Bayer 2014, Yates 2013, Bank 2012,

Bulle 2011, Waldow et al. 2014, Grek 2009, Simola 2005, Sahlberg 2011, Kummulainen & Lankinen 2012) do not present the researchers' own empirical analysis but only refer to the reports or statistics published by the OECD. In the papers where the researchers' own empirical models are being derived and analyzed (e.g., Skryabin et al. 2015, Kriegbaum et al. 2015, Erdogdu & Erdogdu 2015, Tømte & Hatlevik 2011, Zhong 2011, Fonseca et al. 2011), the missing data is most often completely removed, and the sample is analyzed by ignoring the weights and, hence, the population level. Moreover, typically students from only a few countries are being compared in the existing literature, although a very scarce pool exists of comparisons at the level of the whole PISA sample (e.g., Drabowicz 2014, Zhong 2011).

We have also carefully assessed the use of clustering with the PISA data sets and have only been able to identify our own recent publications for the PISA 2012 (Saarela & Kärkkäinen 2014, Saarela & Kärkkäinen 2015^{b,c}) and two older publications for the PISA 2003 (Olsen 2005^b) and for the PISA 2000 (Kjærnsli & Lie 2004). Thus, our main contributions here are that we augment the traditional PISA analysis by utilizing big data LA methods and work with the data set on the macro level of the whole student population, as conforms to the recommendations given by the OECD (2014^b). This population-level scope is a novel setting in big data LA.

8.3 The PISA Profile

In this section, we outline the contextually related work of the chapter. More precisely, since Finland is the primary interest in our clustering application, we introduce the main characteristics of the Finnish educational system, which has performed so well in the PISA assessments, as well as related research. The last part of this section is devoted to a description of the collection and overall processing of the PISA assessment, yielding to multiple forms of publicly available educational data sets on a macro level.

8.3.1 *The Finnish Educational System and the PISA*

In this paper, our main focus is on Finland in comparison to the other countries that participated in the latest PISA assessment. Traditionally, Finnish students have performed exceptionally well in the PISA tests. The reasons for Finland's success on the PISA, particularly in the 2003 and 2006 assessment cycles, have been analyzed in several studies, and educational stakeholders from all over the world have visited Finland to find explanations for the high-performing students.

Consequently, education became an important asset in Finland's image and identity. In fact, Finland has invested considerably in the international educational

export sector (Schatz et al. 2016), and, although Finland's place in the international ranking dropped in the latest PISA assessment, it is still placed the highest in Europe. Here, our goal is to assess the variables that most distinguish Finland from the other countries participating in the PISA.

Finland's high performance in the PISA assessments has been analyzed in several articles. Many of these articles have linked the well-performing students to the highly qualified teachers, who need to have a Master's degree for a permanent position. In particular, it has been argued that, in Finland, being a teacher is one of the most prestigious occupations, as evidenced by the fact that only the best and most motivated students are admitted to the teacher training programs as well as the observation that Finnish teachers enjoy a very high status in society (Morgan 2014, Sahlberg 2011, Linnakylä et al. 2011, OECD 2011, Andere 2015).

A second reason that has been identified as contributing to Finland's high results in the PISA relates to the organization of the national school system. Instead of (i) market-oriented schooling, (ii) standardization of schools and tests, concentrating on measurable performance, and (iii) competition between students and schools, the focus in Finland's schools is more on cooperation, collaboration, and the belief that teachers will support each student's individual learning (Simola 2005, Sahlberg 2011). National curricula as well as explicit learning objectives and standards do exist, but schools and teachers in Finland enjoy great autonomy and decision-making authority (i.e., they can decide on individualized learning strategies and pedagogical methods in order to reach the common educational goals) (Kumpulainen & Lankinen 2012, Linnakylä et al. 2011, OECD 2011).

The fact that schools in Finland are neither competing nor evaluated by standardized tests is one of the reasons why the variance between the Finnish schools is so small⁴ (Simola 2005). Additionally, there is a no division of students into different school types or tracks based on their performance. Indeed, all students in Finland attend common, untracked, comprehensive schools of equally good quality from grades 1–9, typically those nearest to their homes. These schools are publicly funded and offer free lunches, health care, and school transport for all pupils (OECD 2011, Linnakylä et al. 2011).

These mutually interdependent and interconnected factors that are associated with Finland's high achievements in the PISA have also been emphasized by Välijärvi et al. (2007), who have concluded that Finland's success can be explained by a combination of "comprehensive pedagogy, students' own interests and leisure activities, the structure of the education system, teacher education, school practices and, in the end, Finnish culture" (see Table 8.1).

Research has shown that culture tends to affect both people's goals and their actions to reach these goals (Hitlin & Piliavin 2004). As has been pointed out above, Finnish people put great emphasis on equity and equality. Several studies have also highlighted the trust that seems to exist in Finnish culture in general and

⁴ According to the 2012 assessment, the between-school variation in Finland is only 6% of the overall math performance, which is the second-lowest figure in comparison with all PISA countries.

between the educators and the community in particular (Sahlberg 2011, OECD 2011).

Table 8.1 Interaction between culture and education in Finland.

Culture	Education
strong mutual trust	parents and government trust teachers (indicated by the strong autonomy and authority of the teachers)
equity and equality (care for others instead of wanting to be the best)	common untracked comprehensive school systems; free lunch, health care, and school transport; children with special needs study in the same classroom
indulgent country	minimal time allocated to studying, broad rich curriculum

The *Hofstede model* (Hofstede 2011) acknowledges the idea that Finland is more of a collaborative than a competitive country. According to the model, Finland's society can be characterized as being highly "feminine," meaning that the most important driving factors in life are to live a good life and to care for others instead of to focus on one's own success and want to be the best. This is interesting when linked to the recent study by French et al. (2015), who found a negative causal relationship between education expenditure and power distance and masculinity. According to this study, the less masculine a country is, the more it invests in education.

8.3.2 Characteristics and Forms of the PISA Data

The OECD states that the PISA results have a high degree of validity and reliability (for example, OECD 2014b, 2012), so they can be used to assess and compare the educational systems of the participating countries. To ensure the validity and reliability of the PISA data, large amounts of money are spent. For example, in Germany alone, the aggregate costs of the PISA assessment have reached 21.5 million euros (Musik 2016). However, as was pointed out in the introduction of this chapter, the PISA assessments as well as the resulting PISA data are methodologically very complex.

As highlighted by the OECD (2012), "the successful implementation of PISA depends on the use, and sometimes further development, of state-of-the-art methodologies and technologies." Since a mixture of different methods is used in this large study, and many variables are derived, it is not obvious how certain values in the publicly available database⁵ (see Fig. 8.2) were collected, obtained, and report-

⁵ Can be downloaded from <http://pisa2012.acer.edu.au/downloads.php>

ed. The fact that the PISA data are voluminous and complex can also be concluded based on the time that is needed to publish the PISA data and results: Usually around 1.5 years passes between data collection and when the first PISA results and data are published. For example, the 2012 PISA data collection took place in spring 2012, and its results were published in December 2013.

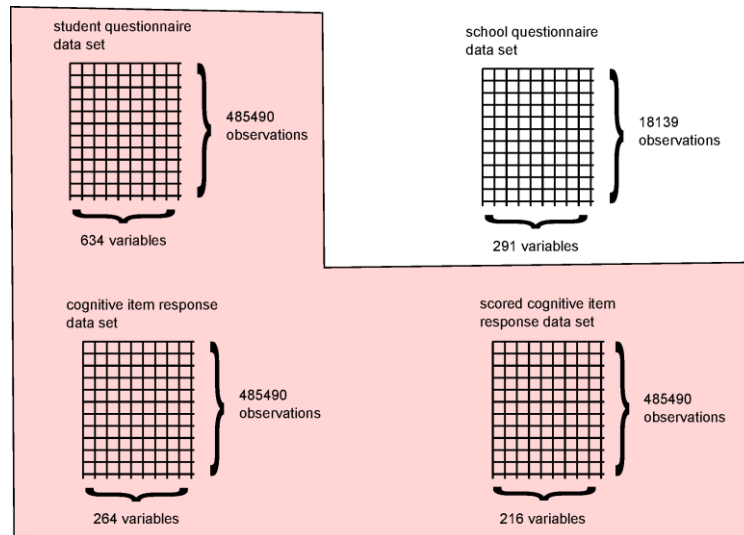


Fig. 8.2 Overview of the 2012 data sets available from OECD.

An overview of the 2012 PISA data is provided in Fig. 8.2. In all three data sets with pink backgrounds in Fig. 8.2, the observations are the assessed students. The basic information about the student (student's ID, country, test language, and school ID) and which test he or she was administered (booklet ID) is provided in all three of these student data sets. The *student cognitive items* and *scored cognitive item response* data sets document the students' responses to the cognitive items and how these were scored. Altogether, there were 206 different cognitive items in the PISA 2012 data. An example of a cognitive item variable label is "SCIE—P2006 Wild Oat Grass Q4." As can be seen, it includes the domain (in this case, science), the PISA cycle in which the question was first used (the PISA 2006), the name for the particular task unit⁶ (*Wild Oat Grass*), and the question number (4).

The most informative and meaningful part of the PISA data is the *student questionnaire data set* (see Fig. 8.2). However, as previously mentioned, one of the biggest challenges when working with the PISA data is that many variables in this data set are not direct measurements but rather variables that have already been transformed and preprocessed. For example, the students' abilities/performances in the cognitive tests are summarized in the form of *plausible values*. Plausible

⁶ PISA items are organized into units. Each unit consists of a stimulus (consisting of a piece of text or related texts, pictures, or graphs) followed by one or more questions.

values are, as Wu (2005) puts it, “multiple imputations of the unobservable latent achievement for each student.” This is explained more thoroughly at the end of this section.

Certain scale indices in the data—indicating, for example, students’ attitudes toward school and learning—are also derived variables. This means that in order to be able to work with the PISA data, it is necessary to understand how the many derived variables have been created and how they can be used for further analysis. In the PISA, the *Rasch model*, which is a special case of item response theory, is used for this purpose.

Gray et al. (2014) emphasize the importance of integrating item response theory factors and methods, such as the Rasch model, into the existing LA models. Item response theory models can improve existing models, because they can model latent (i.e., not directly measurable) traits, such as intelligence, ability, or motivation. Moreover, they can be applied even with a large number of missing values. The potential of using item response theory in LA has been shown, for example, by Bergner et al. (2015), who estimated student abilities based on homework scores from an MOOC in which a large number of scores were missing.

The second challenge when working with the PISA data is the high sparsity. Since the assessment material developed for the PISA exceeds the time that is allocated for the test, each student is administered only a fraction of the whole cognitive testing material and only one of the three different background questionnaires. Because of this rotated design, very few variables in the PISA data sets have values for all observations. For example, in the PISA 2012, each student was assigned a test booklet of cognitive items that should be solvable in two hours. However, the comprehensive PISA 2012 cognitive item battery consisted of test items to be solved in six hours.

The *scored item set* (see Fig. 8.2) incorporates 206 scored items for 485,490 students. Nevertheless, because of the different booklets, which always contain only a fraction of the total items, 74% (that is, 73,860,420) of the different item variables have missing values. Similarly, because of the three different background questionnaires administered, the majority of the variables in the *student questionnaire data set* are missing approximately one-third of their values. We have discussed sparsity in educational data, particularly in the PISA data, and algorithms to cope with this issue in many of our recent studies (Saarela & Kärkkäinen 2014, Saarela & Kärkkäinen 2015a,b,c, Kärkkäinen & Saarela 2015, Saarela et al. 2016b).

Finally, the PISA data are an important example of a large data set that includes weights. Only a fraction of the 15-year-old students from each country takes part in the assessment, but the gathered sample depicts the whole student population by multiplying the students’ results by their respective weights, which simply measure how many similar students are represented by one student in the sample. For example, the sample data of the latest assessment consist of 485,490 students, which, when taking the weights into account, are representative of more than 24 million 15-year-old students in the 68 different countries and territories that participated in the PISA 2012.

Both over- and under-sampling has taken place in the PISA for different student groups. As a consequence, in order to state findings that are valid for the

whole population, it is important to utilize these weights at each stage of the analysis. The way in which we incorporated the weights into a robust clustering algorithm for sparse data is illustrated and applied in our prior works (respectively, Saarela & Kärkkäinen 2015*c,b*).

8.3.3 Rasch Model

As described above, because of the different PISA test booklets administered, the actual scored student test data is extremely sparse with a great deal of missing values (74%). The easiest approach for measuring each student's ability would be to average the percentage of the correct answers over the three domains. However, since not all students were presented with the same test items, and the test items varied in their difficulty, this approach is considered unreliable. With the Rasch model, however, the probability of success on a given item can be modeled as a logistic function of the difference between the student and item parameters (Rasch 1960). Hence, the Rasch model enables a comparison of student abilities/test results/characteristics, even if not all students were tested on the same test items.

In the PISA, the Rasch model is employed to estimate both student abilities—depending on their item responses and the item difficulties in the cognitive test—and general student characteristics—depending on their responses on the background questionnaire. Mathematically, in the simplest case of the Rasch model when the test item is dichotomous, the probability that a student i with ability denoted by β_i provides a correct answer to an item j of difficulty δ_j can be stated as follows (8.1):

$$P(X_{ij} | \beta_i, \delta_j) = \frac{\exp(\beta_i - \delta_j)}{1 + \exp(\beta_i - \delta_j)}, \quad (8.1)$$

When the Rasch model is employed, it iteratively creates a continuum/scale on which both a student's ability and item difficulty are located and where a probabilistic function links these two components. Usually, the item difficulties are estimated first, and this is referred to as the item calibration. The overall objective is to obtain data that will fit the model.

There should be a higher probability that a student should give a correct answer to an easy item than to a difficult item. Similarly, there should be a higher probability that a student with high ability should give correct answers to items than a student with low ability. This is shown in Fig. 8.3, where the probability that a correct answer is given to an item with difficulty $\delta = 0.6$ is plotted for different student abilities. Moreover, as also illustrated in Fig. 8.3, when a student's ability is equal to the difficulty of the item, there is by definition a 50% chance of a correct response in the Rasch model.

To estimate the item difficulty, only the probability of being correct on that item and the ability of the students who completed the item must be known. Likewise, to estimate the student's ability, only the probability of being correct on a set of items and the difficulty of those items must be known (Embretson & Reise 2013). Every item and every student will be located in the scale created with the Rasch model. Therefore, comparable student ability estimates can be obtained, even if the students were assessed with a different subset of items (OECD 2014b). The only requirement is that some link items exist (i.e., some items in the different test booklets must be the same).

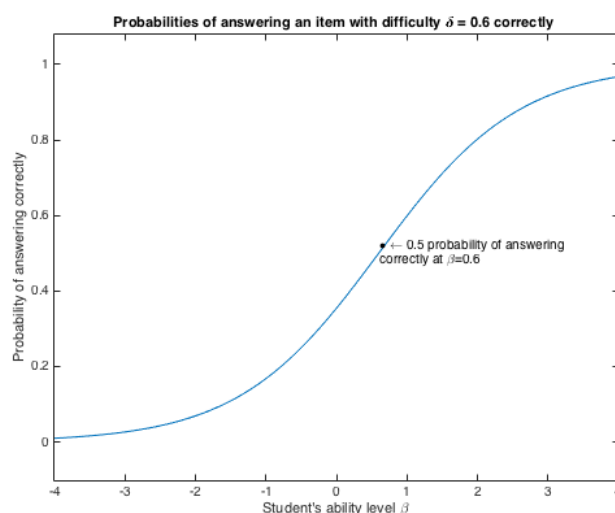


Fig. 8.3. Rasch model example. Probabilities that a correct answer is given to an item with difficulty $\delta = 0.6$ for different student abilities. The probability that a student with ability $\beta = 0.6$ will provide a correct answer to this item is 0.5.

In the PISA, a generalization of the original Rasch model is employed that can score not only dichotomous but also polytomous items (e.g., cognitive items can be scaled as *incorrect*, *partially correct*, and *correct* and questionnaire Likert-scale data can be scaled as *completely agree*, *agree*, *neutral*, *disagree*, and *completely disagree*). This model is called the one-parameter logistic model for polytomous items.

8.3.4 Plausible Values

There exist many other international large-scale educational assessment studies such as the PISA, including the *National Assessment of Educational Progress*⁷, the

⁷ nces.ed.gov/nationsreportcard/

*European Survey on Language Competences*⁸, the *Trends in International Mathematics and Science Study*, and the *Progress in International Reading Literacy Study*⁹. The idea behind the PISA and these other assessments is not to measure and report the proficiencies of individual students. Instead, the primary goal is to provide a reliable overview of the proficiencies and national characteristics of the whole population (OECD 2014b, Marsman 2014). This is the main difference between typical micro- or meso-level LA and big data LA for the PISA.

Plausible values are used to estimate the proficiencies of the population, which, in the PISA, comprises all 15-year-old pupils within the participating countries. Some studies (Monseur & Adams 2008, Wu & Adams 2002, OECD 2014b) have shown that plausible values—in comparison to Weighted Likelihood Estimates, which overestimate, and Expected A Posteriori estimators, which underestimate population variances—produce unbiased estimates for population statistics.

In short, plausible values are random draws from the posterior distribution of a student’s ability. These posterior distributions are estimated with a Bayesian approach in combination with the Rasch model. The posterior distribution of a student’s ability β_i , given his or her vector of item responses \mathbf{x}_i and certain additional variables about the student from the background questionnaire (e.g., gender and many others) that are encoded in a vector \mathbf{y}_i , is defined as (8.2):

$$f(\beta|\mathbf{x}_i, \mathbf{y}_i) \propto P(\mathbf{x}_i|\beta, \delta) f(\beta|\boldsymbol{\lambda}, \mathbf{y}_i), \quad (8.2)$$

where $P(\mathbf{x}_i|\beta, \delta)$ denotes a Rasch model given the student’s ability β and the difficulties of the items δ in the test, and $f(\beta|\boldsymbol{\lambda}, \mathbf{y}_i)$ denotes a population model. This population model for a student i is usually estimated with the latent (called latent because the predictor is unobserved) regression model $\beta_i = \mathbf{y}_i^T \boldsymbol{\lambda} + \varepsilon_i$, where $\varepsilon_i = \mathcal{N}(0, \sigma^2)$ (Marsman 2014, OECD 2014b).

In other words, in each country, the student’s abilities are assumed to follow a conditional Gaussian distribution, given \mathbf{y}_i (i.e., the variables from the background questionnaire). This is the prior distribution. Then, the student takes the PISA test. The statistical model (“likelihood”) of the success in the test is a Rasch model, where the probability of success is a logistic function of the unknown but estimated latent ability and the difficulties of the test items (see Equations 8.1 and 8.2).

The estimated posterior distribution of the ability of the student is specific for each student, as each student has different values for background variables and test results. This means that success in the PISA test “corrects” our prior beliefs regarding the student’s ability. If a student successfully solves a difficult item, this indicates higher ability than success on an easy item. However, the student’s exact ability is not known, and it is represented on the population level with five plausi-

⁸ www.surveylang.org/

⁹ <http://timssandpirls.bc.edu/>

ble values that are random realizations based on his or her posterior distribution. For this reason, the official PISA protocol (OECD 2012) requires that the same analysis be repeated five times when analyzing student performance, with one analysis for each plausible value.

8.4 Comparison of Students in PISA 2012 Countries Using Aggregated Hierarchical Clustering

The empirical part of this work is focused on comparing the student characteristics of Finland to those of the other countries that participated in the PISA assessment 2012. This comparison is conducted by utilizing three of the four LA techniques described by Chatti (2012) (see Section 8.2.1): clustering as one of the core *data mining* techniques, *visualization* of the clustering result to illustrate Finland's position in comparison to the other countries, and, finally, *statistical* testing to verify the findings.

8.4.1 Variables for the Clustering

Our overall analysis method is to apply hierarchical clustering on all PISA 2012 countries/economies, to visualize the similarities between the participating countries through a dendrogram, and to conduct different statistical tests on two distinct levels. For this, we first aggregated the entire sample of half a million students in the PISA 2012 into the population level of each country by computing the weighted means of the available data in a country-wise manner. We used all observations in the PISA 2012 data set. All variables in the PISA student data set (and their possible values) can be found in the codebook¹⁰. In Saarela & Kärkkäinen (2014), Saarela & Kärkkäinen (2015*c,b*) and Kärkkäinen & Saarela (2015), we utilized the individual variables on a student level that are known to explain performance in mathematics. Here, we used an extended set of variables, including those that are more on the scale of a classroom (e.g., teacher behavior) or a country (e.g., time of formal instruction in certain school subjects) than on an individual student level.

In Table 8.2, all variables used in this study are listed. All are derived variables constructed with the Rasch model using students' answers to the background questionnaire or other already-derived variables. For example, the first variable, the *index of economic, social, and cultural status*, is constructed using the *highest parental occupation*, the student's *home possessions*, and the *highest parental education*, which themselves are derived variables constructed with the Rasch model (OECD 2014*b*).

¹⁰ Available at http://pisa2012.acer.edu.au/downloads/M_stu_codebook.pdf.

The following five variables (i.e., those with the IDs 2–6 in Table 8.2) are generally associated with performance on a student level, while the next ten variables (IDs 7–16) are all related to attitudes toward mathematics. Since mathematics was the major domain in 2012, attitudes toward this subject received considerable attention in the background questionnaire. Here, we use all ten mathematics indices that together summarize 67 items in the student background questionnaire.

The next five variables in the table (IDs 17–21) are related to how much time students spend studying. Both formal learning time in different subject areas as well as out-of-school study hours are detailed. The last variable, *Age at ISCED 1*, reports the beginning of the systematic education in reading, writing, and mathematics. The last six variables (IDs 22–27) are all on the level of the teacher or teaching method.

Table 8.2 Overview and identification of the derived PISA variables utilized in this study.

PISA variable	ID	PISA variable	ID
Economic, social, and cultural status	1		
Sense of belonging	2	Attitude toward school: learning outcome	3
Attitude toward school: learning activities	4	Perseverance	5
Openness to problem solving	6		
Self-responsibility for failing in math	7	Interest in mathematics	8
Instrumental motivation to learn math	9	Self-efficacy in mathematics	10
Anxiety toward mathematics	11	Self-concept in mathematics	12
Behavior in mathematics	13	Intentions to use mathematics	14
Subjective norms in mathematics	15	Mathematics work ethic	16
Out-of-school study time	17	Learning time (min. per week)—Test language	18
Learning time (min. per week)—Mathematics	19	Learning time (min. per week)—Science	20
Age at <ISCED 1>	21		
Teacher-student relations	22	Mathematics teacher's support	23
Teacher behavior: formative assessment	24	Teacher behavior: Student orientation	25
Teacher behavior: Teacher-directed instruction	26	Experience with applied math tasks at school	27

8.4.2 Hierarchical Clustering

An issue with the PSA data is the aforementioned absence of a large number of values. Moreover, each student in the PISA data sets has a weight expressing how representative he or she is of the population of all 15-year-old students within his or her country. Therefore, we computed the weighted means of the available data for each variable for each country/economy as inputs for the clustering algorithm.

We then normalized our data set using z-scoring and applied hierarchical clustering with Matlab's default settings (i.e., agglomerative single-linkage clustering with the Euclidean distance).

Agglomerative clustering techniques operate in a bottom-up fashion (Zaki & Meira Jr 2014). Hence, we started with each PISA country as a separate cluster. Then, the most similar country clusters C_m and C_n were repeatedly merged so that they formed a new and bigger cluster. The most similar clusters were defined as the ones with the smallest Euclidean distance between a point in C_m and a point in C_n (8.3):

$$\delta(C_m, C_n) = \min\{\delta(u, v) | u \in C_m, v \in C_n\}, \quad (8.3)$$

where $\delta(u, v) = \left(\sum_{i=1}^d (u_i - v_i)^2\right)^{\frac{1}{2}}$ (see Zaki and Meira Jr 2014).

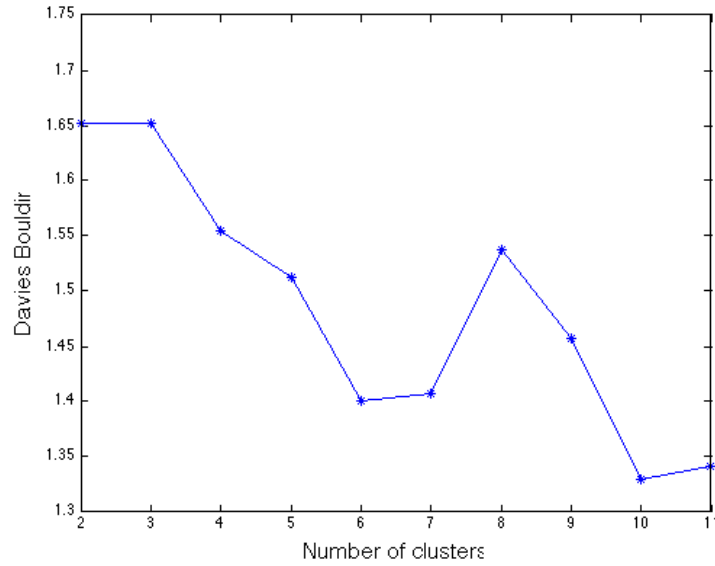


Fig. 8.4 The Davies-Bouldin index suggests that there are ten clusters in the data.

To decide the number of clusters in the PISA 2012, the Davies-Bouldin cluster index (Davies & Bouldin 1979) was applied on the z-scored data. As can be seen from Fig. 8.4, the Davies-Bouldin index suggested that there are ten clusters in the data. Therefore, the merging of closest clusters was terminated after ten clusters were formed.

8.5 Results

In this section, we first visualize the hierarchical clustering result of the aggregated PISA countries in the form of a dendrogram. Then, we profile the country clusters according to their geographic and cultural similarities. Finally, we analyze the clustering results more deeply using statistical tests on two different levels. Since Finland is our primary interest, we first evaluate the differences between all clusters, and then we analyze Finland's cluster and its position within its own cluster.

8.5.1 Visualization and Profiling of the Clusters

Fig. 8.5 shows the hierarchical clustering result. Based on the similarities of countries in particular groups, we suggest the following labels for the ten clusters, as documented in Table 8.3.

Table 8.3 Clustering results.

ID	Label	Countries/economies
C1	"Nordic/English-speaking"	Australia, Canada, United Kingdom, New Zealand, Florida (USA), Connecticut (USA), Massachusetts (USA), USA, Denmark, Iceland, Norway, Sweden
C2	-	Costa Rica, Israel, Uruguay
C3	"Eastern countries"	Bulgaria, Lithuania, Montenegro, Perm (Russia), Romania, Russia, Serbia
C4	"South America/Africa"	Argentina, Chile, Tunisia
C5	"developing countries"	Brazil, Colombia, Indonesia, Mexico, Malaysia, Peru, Thailand, Turkey, Vietnam
C6	"high-performing Asian"	Shanghai (China), Singapore
C7	"Kazakhstan"	Kazakhstan
C8	"Arabic"	United Arab Emirates, Jordan, Qatar
C9	"Asian"	Hong Kong (China), Japan, Korea, (Macao) China, Taiwan
C10	"Europe"	Austria, Belgium, Switzerland, Czech Republic, Germany, Spain, Estonia, Finland, France, Greece, Croatia, Hungary, Ireland, Italy, Liechtenstein, Luxembourg, Latvia, Netherlands, Poland, Slovak Republic, Slovenia

It is a surprise that Finland is not part of the Nordic/English-speaking cluster to which all other Nordic countries belong. This finding is interesting compared to the classification of Bulle (2011), who introduces "the Northern model: Denmark, Finland, Iceland, Norway, Sweden" as one of the five main OECD educational systems. This indicates that even if the educational systems are similar, it does not necessarily follow that the student characteristics are also similar.

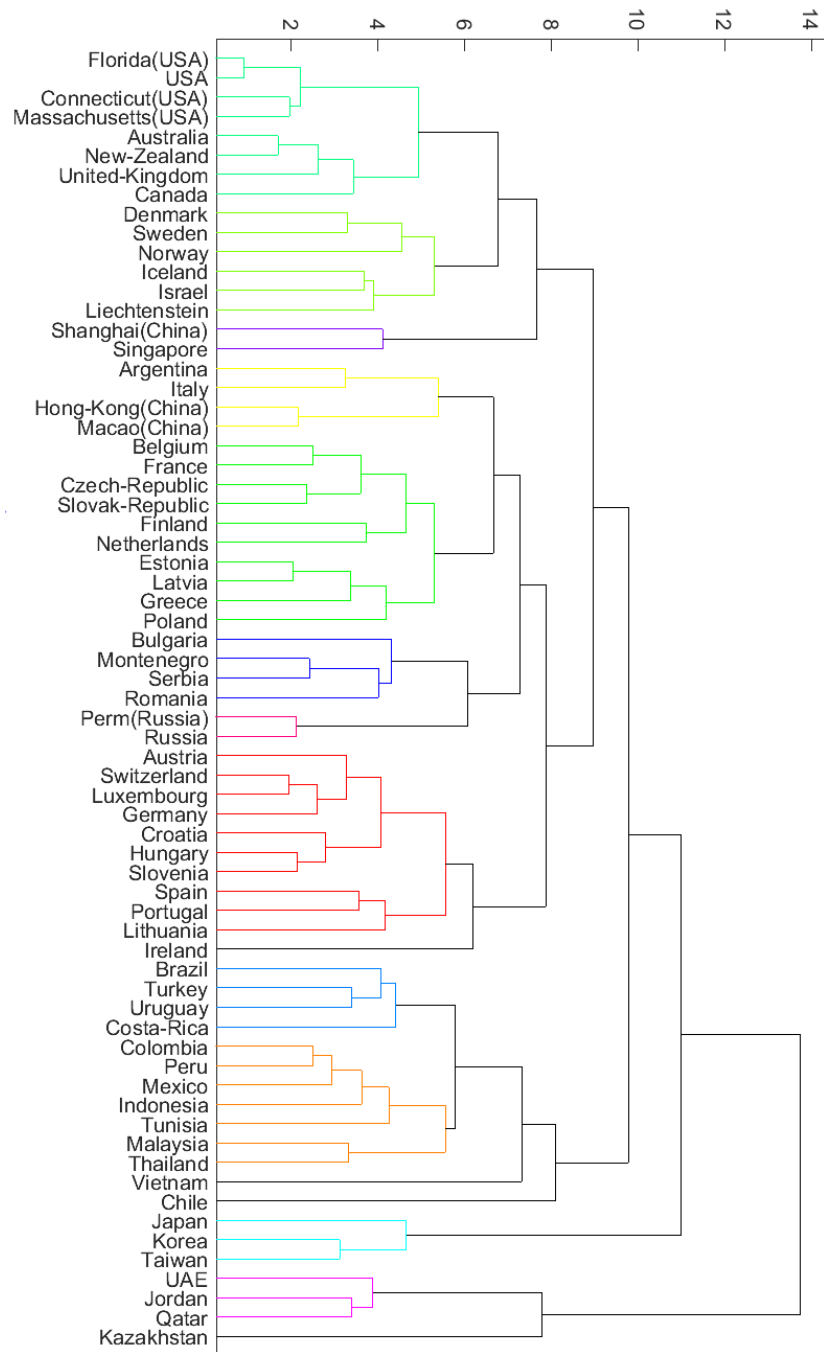


Fig. 8.5 Dendrogram of all countries when their weighted mean is clustered.

The dendrogram implies that Finland belongs to the Europe cluster and is actually closest to the Netherlands. In the PISA 2012 results summary (OECD 2014a, page 7), the performances of these two countries in mathematics were found to not be statistically significantly different among many other pairs of countries. In addition, both the Netherlands and Finland are highly feminine cultures according to the *Hofstede model* (Hofstede 2011).

As has been explained above, it was unexpected that Finland belonged to the Europe cluster and not to the Nordic/English-speaking cluster. We utilized statistical tests to assess the significance of the single variables and to explain why a particular country was allocated to a certain cluster. Since not all of our variables were normally distributed, we had to use non-parametric tests.

To specifically address the finding of Finland's position, we will first report the differences between all the clusters. Second, we will summarize the differences between Finland and its own Europe cluster; third, we will describe the variables that separate the Europe cluster from the Nordic/English-speaking cluster.

8.5.2 Differences Between All the Global Clusters

A Kruskal-Wallis H test (Kruskal & Wallis 1952) showed that there was a highly statistically significant difference in 20 of the 27 variables between the different clusters. The test statistics of all highly statistically significant variables are provided in Table 8.4. With reference to Table 8.4, variable 25, *teacher behavior: student orientation* (i.e., how much attention that teachers pay to individual students), was the most important in terms of accounting for variance in the cluster membership ($\chi^2(9) = 51,227, p < 0.001$).

Table 8.4 Kruskal-Wallis H test statistics (all clusters) with a post hoc test.

variable	$\chi^2(9)$	p	Post hoc test	variable	$\chi^2(9)$	p	Post hoc test
1	48,676	***	C10-C5, C1-C5	4	38,499	***	C9-C1
5	33,306	***	-	7	37,399	***	-
8	48,701	***	C10-C5	9	49,857	***	C9-C5, C10-C5
10	30,765	***	-	11	42,170	***	C1-C5
12	35,298	***	-	13	49,549	***	C1-C5
14	34,029	***	-	15	49,082	***	C10-C5
16	39,863	***	-	18	40,457	***	-
19	36,542	***	-	22	42,940	***	C10-C5
23	46,378	***	C10-C5	24	45,203	***	-
25	51,227	***	C10-C5	26	42,610	***	-

Subsequently, pairwise comparisons were performed using Dunn's (1964) procedure with a Bonferroni correction for multiple comparisons. This post hoc anal-

ysis revealed highly statistically significant differences in the *ESCS* between the developing (mean rank = 5.67) and the Nordic/English-speaking clusters (mean rank = 57.25) as well as between the developing and the Europe (mean rank = 40.47) clusters, but not between any other group combination for this variable. This is also illustrated in Fig. 8.6, in which all pairwise comparisons of the different clusters for their *ESCS* are shown. In the figure, black lines reflect a pairwise comparison that is not statistically significant, while orange lines reflect a statistically significant pairwise comparison.

The last column in Table 8.4 summarizes the post hoc analysis for all the variables. As can be seen from the table, highly statistically significant differences were found in the *attitude toward school: learning activities* (i.e., the degree to which a student sees hard work in school pay off later) between the Asian (mean rank = 5.00) and the Nordic/English-speaking clusters (mean rank = 51.08), in the *interest in and enjoyment of mathematics* between the developing countries (mean rank = 56.89) and Europe (mean rank = 14.90) clusters, in the *instrumental motivation to learn mathematics* (i.e., the degree to which a student's hard work in mathematics pays off later) between the developing (mean rank = 57.89) and the Asian (mean rank = 7.80) countries, and between the developing countries and the Europe (mean rank = 19.10) clusters.

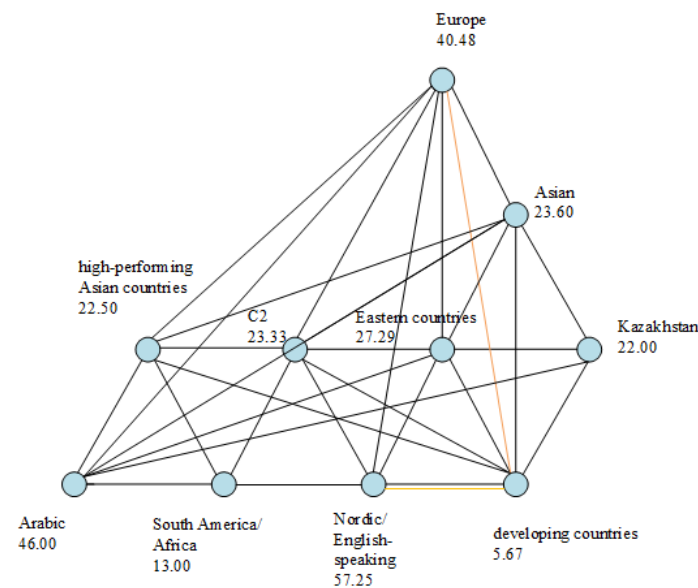


Fig. 8.6 Pairwise comparisons of clusters for *ESCS*. Statistically significant differences (between developing countries and the Nordic/English-speaking cluster and between developing countries and the Europe cluster) are marked in yellow.

Highly statistically significant differences were found for the developing countries cluster when compared with the Nordic/English-speaking cluster with regard to *anxiety toward mathematics* (mean rank C5 = 55.00 vs. C1 = 14.92) and *behav-*

ior in mathematics (i.e., the role of mathematics inside and outside school) (mean rank C5 = 54.11 vs. C1 = 12.17). In addition, highly statistically significant differences were found for the developing countries cluster when compared with the Europe cluster with regard to *subjective norms in mathematics* (mean rank C5 = 51.11 vs. C10 = 15.81) (i.e. how much attention to mathematics is given by friends and family), *teacher-student relations* (mean rank C5 = 51.44 vs. C10 = 14.90), *mathematics teacher's support* (mean rank C5 = 52.22 vs. C10 = 14.43), and *teacher behavior: student orientation* (mean rank C5 = 54.33 vs. C10 = 15.14), respectively. No highly statistically significant differences were found for any other group combination.

Hence, the statistical test on a global level suggests that, overall, the Europe cluster and the developing countries cluster are the most dissimilar to each other. Students in the Europe cluster have a higher economic, social, and cultural status—but students in the developing countries cluster have higher interests, more motivation to learn, and higher subjective norms in mathematics from their friends and family. Furthermore, students in the developing countries tend to report better relations with their teachers.

When comparing Finland to other countries, the rather negative attitudes toward mathematics were already observed in the 2003 assessment cycle. In both *interest in* and *enjoyment of* mathematics, Finland was ranked 37th out of the 40 participating countries (Linnakylä et al. 2011).

Moreover, in a longitudinal study of Finnish students in grade 1 to grade 12 by Metsämuuronen et al. (2012), it was concluded that student contentment in regard to school in Finland decreases significantly from the second to the eighth grade, while it then very slightly increases starting in the ninth grade. The majority (82%¹¹) of the Finnish students participating in the PISA are in the ninth grade, and almost all the rest are in the eighth grade (16%). Hence, Finnish students are at the stage in their basic education where their self-reported attitudes toward school are very poor.

Metsämuuronen et al. (2012) suggest that these generally negative attitudes of Finnish students toward education are due to their modesty and honesty: “Part of the explanation in Finland [...] can be the appreciation of honesty and speaking frankly [...] pupils in Finland [...] are relatively humble when they describe their knowledge. This ‘humbleness’ may also be reflected in attitude measurements.”

8.5.3 Differences Between Finland and the Other Countries Within the Europe Cluster

According to the clustering result, Finland is most similar to the countries in the Europe cluster. Table 8.5 summarizes the highly statistically significant variables according to which Finland differs from the remaining countries within its own cluster, as determined by the Wilcoxon signed-rank tests.

¹¹ Our own calculation from the PISA 2012 data.

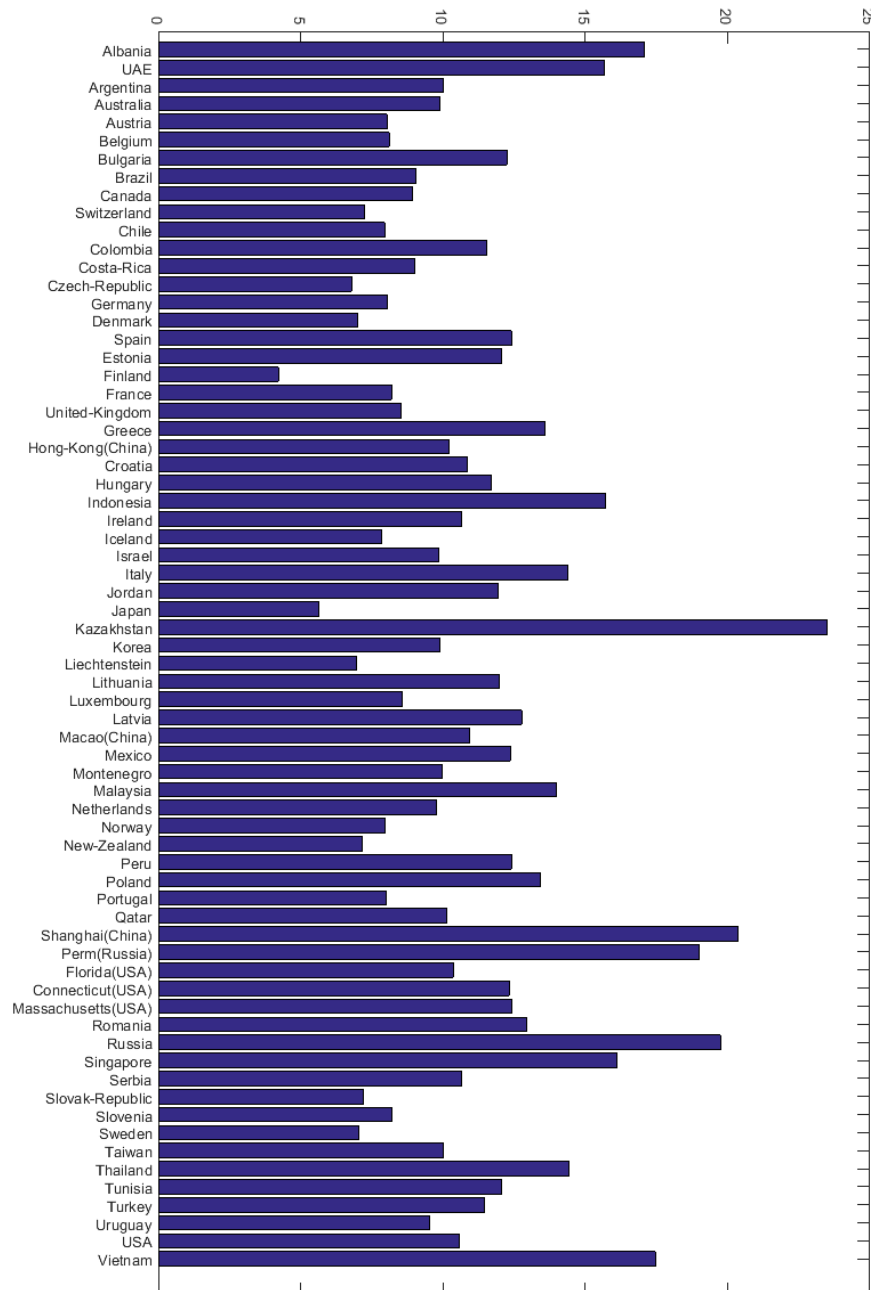


Fig. 8.7 Weighted averages of the out-of-school study hours for all in PISA-participating countries. In comparison to all the other countries, Finnish students study the least after school.

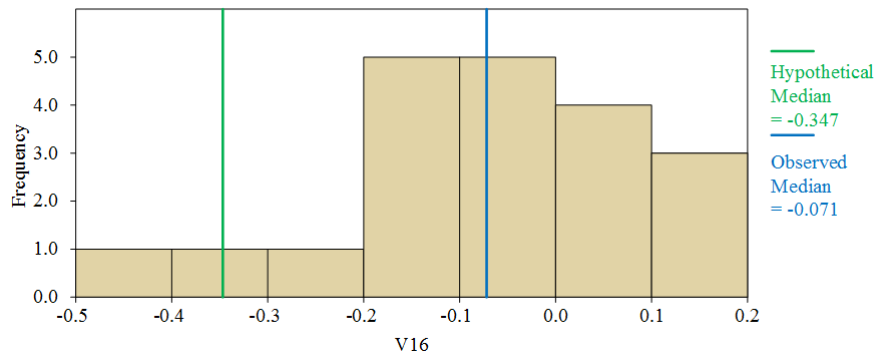
Table 8.5 Wilcoxon signed-rank statistics (Europe - Finland clusters).

variable	1	7	10	11	16	17	18	24
Z	-3.920	3.920	3.920	3.771	3.808	3.920	3.845	3.920
P	***	***	***	***	***	***	***	***

As can be seen in Table 8.5, the majority of the Europe cluster has a significantly lower *ESCS* than Finland ($z = -3.92$, $p < 0.001$). Nevertheless, the Europe cluster majority has a significantly higher *self-responsibility for failing in mathematics* ($z = 3.92$, $p < 0.001$), *anxiety toward mathematics* ($z = 3.771$, $p < 0.001$), and *self-efficacy in mathematics* ($z = 3.92$, $p < 0.001$) than Finland. Furthermore, the Europe cluster in general shows higher scores in many variables that measure emphasis of formal assessment and how much time students spend studying.

In particular, there is a significantly higher *work ethic in mathematics* ($z = 3.808$, $p < 0.001$) and more *out-of-school study hours* in the Europe cluster than in Finland ($z = 3.920$, $p < 0.001$). The latter is illustrated in Fig. 8.7, where the weighted average out-of-school study hours for students in all participating PISA countries are plotted. As can be seen from the figure, Finnish students not only study the least outside of school within their own Europe cluster but also compared to all other countries participating in the PISA.

In addition, *learning time (min. per week) - test language* in Europe is significantly greater than in Finland ($z = 3.845$, $p < 0.001$, see Table 8.5), and Europe has a significantly higher score in *teacher behavior: formative assessment* than Finland ($z = 3.920$, $p < 0.001$). In summary, these results support the observations by Sahlberg (2011), who writes that educational decision makers in Finland “do not seem to believe that doing more of the same in education would necessarily make any significant difference for improvement.”

**Fig. 8.8** One-Sample Wilcoxon Rank Test for work ethic: The work ethic of students in Finland is significantly lower than the work ethic of students in the Europe cluster.

As can be seen from the Wilcoxon signed-rank test result and Fig 8.8, 15-year-old students in Finland seem to already have a rather relaxed attitude toward for-

mal assessment and investing time in their studies. This is particularly evident in the highly statistically significantly lower work ethic¹² of Finnish students.

It must also be kept in mind that the systematic teaching of reading, writing, and mathematics begins later in Finland than in Europe ($z = -3.435$, $p < 0.001$). This is illustrated in Fig. 8.9. In Finland, children are seven years old when they start school. Combined with the finding that the hours of formal instruction of certain subjects are, as described in the above paragraph, significantly lower in Finland, this means that Finnish students spend less time at school than students in other countries. This finding has also been emphasized by Kumpulainen & Lankinen (2012).

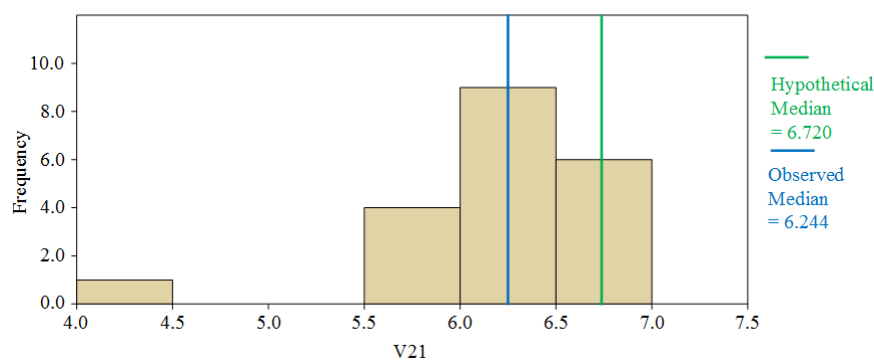


Fig. 8.9 One-Sample Wilcoxon Rank Test for *age at <ISCED 1>*: Systematic teaching of reading, writing, and mathematics begins significantly later in Finland than in Europe.

8.5.4 Europe Cluster in Comparison to the Nordic/English-Speaking Cluster

A Mann-Whitney U test was run to determine if there were differences in the 27 variables between the Europe and the Nordic/English-speaking clusters. Distributions of the 27 variables for the two groups were not similar, as assessed by visual inspection. The test statistics can be found in Table 8.6.

When we combine the test results of the Mann-Whitney U test of the Nordic/English-speaking versus Europe and the Wilcoxon signed-rank test of Europe versus Finland, we find that two variables (16 and 18) augment Finland's special

¹² The *work ethics* scale index is computed with the Rasch model and by using the extent to which students agree or disagree with the following statements: *I finish my homework in time for mathematics class; I work hard on my mathematics homework; I am prepared for my mathematics exams; I study hard for mathematics quizzes; I keep studying until I understand mathematics material; I pay attention in mathematics class; I listen in mathematics class; I avoid distractions when I am studying mathematics; and I keep my mathematics work well organized.*

characteristics: *work ethic* and *study time (test language)* are statistically significantly lower in Europe and even lower in Finland. As described above, these variables measure how much time students spend studying and how much they strive for high grades in mathematics.

Table 8.6 Mann-Whitney U test results comparing the Europe cluster to the Nordic/English-speaking cluster.

PISA variable ID	4	8	9	12	15	16	18	22	23	25
U	19	27	5	30	1	38	22	20	28	20
Z	-4.004	-3.705	-4.528	-3.593	-4.678	-3.293	-3.892	-3.967	-3.668	-3.967
p	***	***	***	***	***	***	***	***	***	***

According to the Mann-Whitney U test, there was a significant ($p < 0.001$) difference in *attitude toward school: learning activities, interest in and enjoyment of mathematics, instrumental motivation to learn mathematics, self-concept in mathematics, subjective norms in mathematics, mathematics work ethic, test language learning time, teacher-student relations, mathematics teacher's support, and teacher behavior: student orientation* between the two clusters. In all these variables, the Nordic/English-speaking cluster showed higher values than the Europe cluster. With reference to Table 8.6, *subjective norms in mathematics* seems to be the most important variable that separates the Nordic/English-speaking cluster from the Europe cluster.

The comparisons of the Nordic/English-speaking cluster to the Europe cluster mostly revealed variables that estimate the students' own perception of their merits and importance. It is especially interesting that the self-reported self-concept is significantly lower in Finland, because this PISA 2012 variable actually explains the performance of Finnish students in the PISA mathematics test fairly well, and it is the mathematics scale index that correlates the most with their plausible values in mathematics (Saarela & Kärkkäinen 2014). However, it seems that even if Finnish students evaluate their own skills realistically, they are more modest about them. Generally, students in the Nordic/English-speaking cluster tend to have higher opinions about themselves, are more motivated, and report better relations with their teachers.

The average mathematics performance based on the plausible values of the countries in the Nordic/English-speaking cluster is 495.3, while the mean mathematics performance of the countries in the Europe cluster is higher (500.5). We conclude that learning time and positive student-teacher relations seem to be less important features than collaborative skills or being free from arrogance for explaining students' success in the PISA test.

8.6 Visual LA of the PISA Results

The macro-level LA of the Finnish basic educational system is visualized in the dashboard of Fig. 8.10–8.13 through the lens of the cultural background, the PISA, and our empirical analysis. This dashboard consists of four figures, and its composition was inspired by Ferguson & Shum (2012).

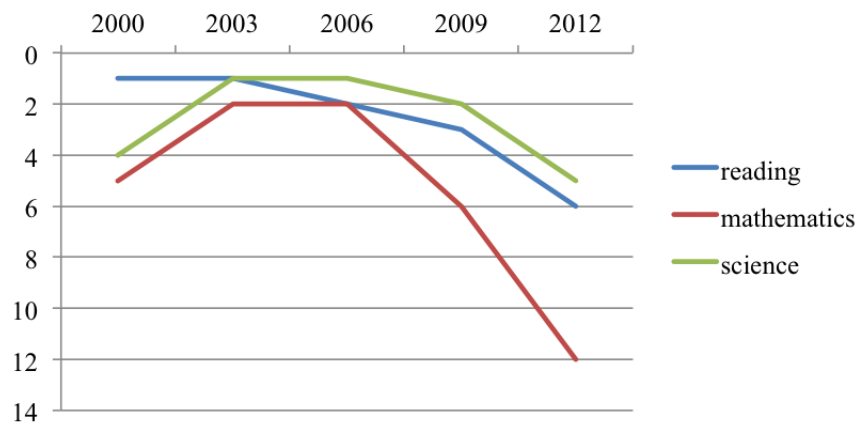


Fig. 8.10. Finland's ranking in the PISA cycles.

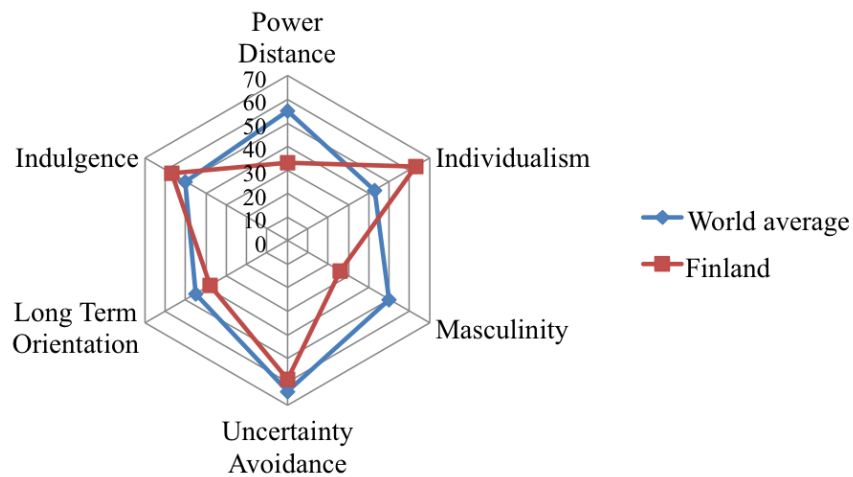


Fig. 8.11. Finland's scores in the Hofstede model dimensions.

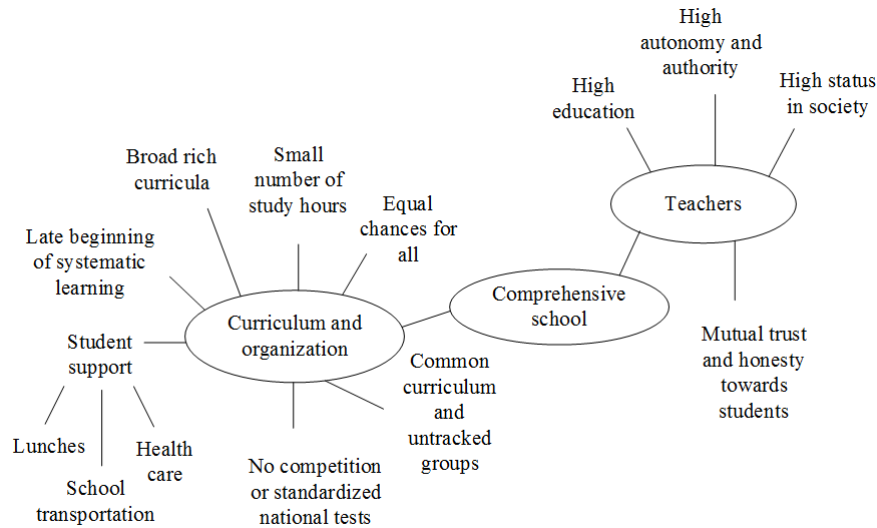


Fig. 8.12. Characteristics of the Finnish educational system.

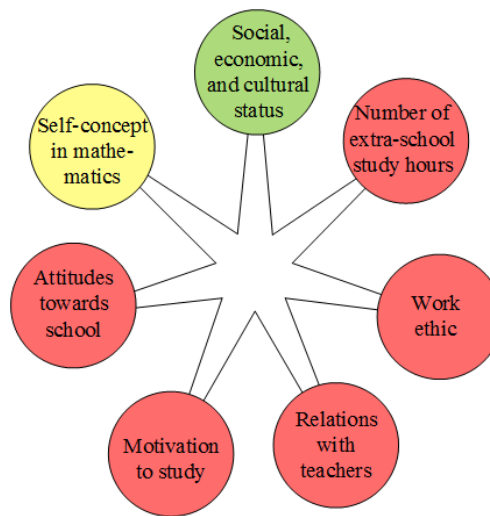


Fig. 8.13. Finland's student characteristics from clustering. The red bubbles indicate alarming characteristics and the yellow bubble indicates the characteristic that could be improved.

Finland has been a top-performing PISA country in the last five assessment cycles (Fig. 8.10), although the ranking clearly decreased in 2012, especially in mathematics. A certain interesting success factor of the educational system is the cultural deviation from the world's midlevel as a feminine culture with a low power distance (Fig. 8.11). The system is based on the strong autonomy and authority of highly educated teachers, with a small amount of formal assessment and,

in particular, a complete lack of national comparative assessments of the learning results (Fig. 8.12). In addition, a rich common curriculum is present for untracked groups of students, who start late in their systematic learning of reading, mathematics, and science. As a whole, equity and equality characterize the system, which provides strong student support (e.g., in the form of free lunches, health care, and school transportation) (Fig. 8.12).

However, many contradictory factors about the Finnish students in relation to their high PISA results emerged in the empirical LA analysis (Fig. 8.13): they have a low motivation to learn and excel in school, a low interest in school topics, a low work ethic, and an exceptionally small number of extra-school study hours. The importance of their studies, and specifically mathematics, is considered low for their future career. The overall evaluation of the different facets of the dashboard indicates that the lowering trend of the PISA, and particularly the mathematics performance of Finnish students, may continue. To improve the system, so as to perhaps be ranked once again as number one in the PISA, students need to be more motivated and oriented toward schoolwork, extra-school study hours, and mathematics, and to keep their future career orientation clearly in mind. We also hypothesize that the complete common, joint, and untracked subject orientations demotivate the most talented students by requiring minimal effort from them. All these factors provide further challenges to subsequent upper secondary and higher education.

8.7 Discussion

We briefly summarize the empirical findings from the previous sections. These were obtained by utilizing one of the illuminated educational clustering techniques, hierarchical clustering, and by taking into account all the specific demands of the PISA data discussed above. As suggested by the Davies-Bouldin cluster validation index, we first divided the students of all the PISA-participating countries into ten separate groups. The clusters that were found generally could be explained by the culture and geographical location of the countries in them. Nevertheless, Finland surprisingly belonged to the Europe cluster (see Fig. 8.5), while all the other Scandinavian countries belonged to the cluster of Nordic/English-speaking countries. This illustrates how similar educational systems (see Bulle 2011) can be reflected by different student characterizations.

Statistical significance tests of the clustering result revealed why particular countries were allocated to a certain cluster. At first, it seemed that the results of the statistical test were somehow contradictory, as students in better-performing countries had worse student-teacher relations and generally showed less confidence in their own achievements and skills. Moreover, the work ethic of the students in the better-performing Europe cluster was significantly lower than that of the students in the Nordic/English-speaking countries cluster—and the better-performing Finnish students showed a work ethic that was significantly worse than

the remaining students in the Europe cluster. However, these findings seem to be connected to and explicable by the existing research related to Finnish culture in general.

As was explained in the literature review about the Finnish educational system and culture, Finnish citizens are modest about their own achievements, and they place great emphasis on equity and equality. The most important driving factors in the life of this highly feminine country are to live a good life and to care for others rather than to focus on one's own success and desire to be the best. This is interesting because, as emphasized in our literature review, French et al. (2015) found a negative causal relationship between education expenditures and power distance and masculinity. Furthermore, Finnish students seem to have an extremely relaxed attitude toward formal assessment and investing time in studies, as can be expected in a feminine country.

Finally, the main success of Finnish students in the PISA seems to a great extent to be related to the relatively better scores of the lowest-scoring Finnish students in comparison with other countries (Andersen 2010), which in turn is supported by the collaborative and ostentation-free thinking in the country. However, as illustrated in Fig. 8.10, Finland's ranking significantly dropped in the latest PISA 2012 assessment (OECD 2013b), and according to the overall characterization of the Finnish students as given and visualized in Fig. 8.13, the negative trend in performance might have continued in the PISA 2015¹³.

8.8 Conclusions

LA is a growing and expanding research field. Traditionally, many studies have concentrated on analyzing educational data originating from a macro or (at the most) meso level. The publicly available and high-quality PISA data sets, on the other hand, provide the opportunity to conduct big data LA research on the macro level, because they comprise data on a whole population of international students.

In this chapter, we have introduced the background for conducting large-scale LA research on the PISA. We have described the main data sets as well as the complexities within them and discussed how to work with these data. Moreover, we have provided a review of relevant clustering studies within the educational domain. Our empirical work, as discussed in the previous section, provided novel findings and strengthened earlier knowledge on the particularities of the Finnish educational system, which has received a great deal of attention during the 21st century due to the exceptionally good performance of the Finnish students in the PISA tests.

We used quantitative LA methods to identify the main attributes of individual learners that affect their learning experience in the environment where the learning occurs (Fournier et al. 2011). Similar to the reviewed educational clustering studies in Section 8.2.2, we analyzed the student model; however, in contrast to these

¹³ Data from the PISA 2015 will be published by the OECD in December 2016 (National Center for Education Statistics 2016).

previously reviewed studies, our model represented a prototype of a national student population obtained by weighted aggregation. Concerning Finland, the high-achieving country inside the PISA assessments, it was concluded that an educational system promoting student collaboration, humility, and equity can successfully cope with the challenges of negative attitudes toward mathematics, low work ethic, and little study time outside school. This summarizes the evidence-based knowledge discovered about the long-term impact of educational policies and practices on the achievement targets (Piety et al. 2014). Such a conclusion also provides an example of a national education system assessment using big data LA as illustrated in Fig. 8.1: The international-objectives driven data collection and transformation improves understanding of educational arrangements via proper analysis methods that are able to cope with the specialties of the sampled large-scale data.

Big data LA, as described in Section 8.1 and depicted in Fig. 8.1, linking together the four dimensions of LA proposed by Chatti et al. (2014) (see also Greller & Drachslar 2012), encapsulated and supported the overall management of the large-scale educational system assessment based on the PISA data. Our empirical work exemplifies the multiple facets of LA: hierarchical clustering as a data mining technique, visualization of the dendrogram to illustrate the clustering result, and statistical testing to verify the findings. Thus, our work increased the body of knowledge for the macro level of educational systems. We promoted reflection of the main characteristics that differentiate the students in various educational environments, according to the objectives of LA by Chatti et al. (2014) (see Section 3.3). Our reflections of the PISA results were emphasized in the dashboard in Fig. 8.10–8.13 using different LA visualization tools. This dashboard facilitates awareness and monitoring of critical educational aspects for the Finnish 15-year-old student population (Beheshitha et al. 2016).

As a whole, the PISA—as well as the other large-scale-assessments, such as those mentioned in Section 8.3.4—provides a very rich and interesting source for macro-level LA studies. We think that the methods and the framework developed for the publicly available large-scale assessment data sets can and will advance the open architecture of educational applications, which Peña-Ayala (2014) has identified as one of the shortcomings of the current educational data analysis research area.

As part of our future research, we intend to repeat our study using the individual students instead of the country-level aggregation as data for clustering. Furthermore, one of the recent trends in LA focuses on educational process mining (Sedrakyan et al. 2016, Mukala et al. 2015, Trčka et al. 2010). For the traditional pen-and-paper PISA tests, this is not an option. However, for the future PISA cycles, where the tests will be increasingly conducted electronically and log event data will therefore be available (compare the PISA 2012 problem-solving test, which was conducted electronically and log files can be downloaded from the above-cited OECD webpage), this would provide an interesting and promising direction for future research.

Acknowledgements

The authors would like to thank Dr. Jouni Välijärvi, Dr. Kari Nissinen, and Dr. Antero Malin from the Finnish Institute for Educational Research for the valuable discussions and support concerning the PISA data and assessment. We also acknowledge BSc Susanne Jauhiainen for technical assistance in preparing the final version of the chapter.

References

- Allen LK, Mills C, Jacovina ME, Crossley S, D’Mello S, McNamara DS (2016) Investigating Boredom and Engagement During Writing Using Multiple Sources of Information: The Essay, the Writer, and Keystrokes. In: Proceedings of the Sixth International Conference on Learning Analytics & Knowledge, ACM, pp 114–123
- Almeda MV, Scupelli P, Baker RS, Weber M, Fisher A (2014) Clustering of design decisions in classroom visual displays. In: Proceedings of the Fourth International Conference on Learning Analytics & Knowledge, ACM, pp 44–48
- Andere E (2015) Are Teachers Crucial for Academic Achievement? Finland Educational Success in a Comparative Perspective. *Education Policy Analysis Archives* 23(39):1–27
- Andersen FØ (2010) Danish and Finnish PISA results in a comparative, qualitative perspective: How can the stable and distinct differences between the Danish and Finnish PISA results be explained? *Educational Assessment, Evaluation and Accountability* 22(2):159–175
- Auld E, Morris P (2016) PISA, policy and persuasion: translating complex conditions into education ‘best practice’. *Comparative Education* 52(2):202–229
- Bank V (2012) On OECD policies and the pitfalls in economy-driven education: The case of Germany. *Journal of Curriculum Studies* 44(2):193–210
- Beheshitha SS, Hatala M, Gašević D, Joksimović S (2016) The Role of Achievement Goal Orientations When Studying Effect of Learning Analytics Visualizations. In: Proceedings of the Sixth International Conference on Learning Analytics & Knowledge, ACM, pp 54–63
- Bergner Y, Colvin K, Pritchard DE (2015) Estimation of Ability from Homework Items when There Are Missing and/or Multiple Attempts. In: Proceedings of the Fifth International Conference on Learning Analytics & Knowledge, ACM, pp 118–125
- Blikstein P, Worsley M, Piech C, Sahami M, Cooper S, Koller D (2014) Programming Pluralism: Using Learning Analytics to Detect Patterns in the Learning of Computer Programming. *Journal of the Learning Sciences* 23(4):561–599
- Bogarin A, Romero C, Cerezo R, Sanchez-Santillan M (2014) Clustering for Improving Educational Process Mining. In: Proceedings of the Fourth International Conference on Learning Analytics & Knowledge, ACM, pp 11–15
- Bouchet F, Harley JM, Trevors GJ, Azevedo R (2013) Clustering and Profiling Students According to their Interactions with an Intelligent Tutoring System Fostering Self-Regulated Learning. *Journal of Educational Data Mining* 5(1):104–146
- Brown MG, DeMonbrun RM, Lonn S, Aguilar SJ, Teasley SD (2016) What and when: The Role of Course Type and Timing in Students’ Academic Performance. In: Proceedings of the Sixth International Conference on Learning Analytics & Knowledge, ACM, pp 459–468
- Bulle N (2011) Comparing OECD educational models through the prism of PISA. *Comparative Education* 47(4):503–521
- Chandra E, Nandhini K (2010) Knowledge mining from student data. *European Journal of Scientific Research* 47(1):156–163

- Chatti MA, Dyckhoff AL, Schroeder U, Thüs H (2012) A Reference Model for Learning Analytics. *International Journal of Technology Enhanced Learning* 4(5–6):318–331
- Chatti MA, Lukarov V, Thues H, Muslim A, Yousef AMF, Wahid U, Greven C, Chakrabarti A, Schroeder U (2014) Learning Analytics: Challenges and Future Research Directions. *E-learning and Education (Eleed) Journal* 10:1–16
- Coffrin C, Corrin L, de Barba P, Kennedy G (2014) Visualizing patterns of student engagement and performance in MOOCs. In: *Proceedings of the Fourth International Conference on Learning Analytics and Knowledge*, ACM, pp 83–92
- Colthorpe K, Zimbardi K, Ainscough L, Anderson S (2015) Know Thy Student! Combining Learning Analytics and Critical Reflections to Increase Understanding of Students Self-Regulated Learning in an Authentic Setting. *Journal of Learning Analytics* 2(1):134–155
- Davies DL, Bouldin DW (1979) A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-1*(2):224–227
- Dawson S, Gašević D, Siemens G, Joksimović S (2014) Current state and future trends: A citation network analysis of the learning analytics field. In: *Proceedings of the Fourth International Conference on Learning Analytics & Knowledge*, ACM, pp 231–240
- Dawson S, Gašević D, Mirriahi N (2015) Challenging assumptions in learning analytics. *Journal of Learning Analytics* 2(3):1–3
- Deng Z, Gopinathan S (2016) PISA and high-performing education systems: explaining Singapore education success. *Comparative Education* 52(4):449–472
- Desmarais MC, Lemieux F (2013) Clustering and Visualizing Study State Sequences. In: *Proceedings of the 6th International Conference on Educational Data Mining*, pp 224–227
- Drabowicz T (2014) Gender and digital usage inequality among adolescents: A comparative study of 39 countries. *Computers & Education* 74:98 – 111
- Dunn OJ (1964) Multiple comparisons using rank sums. *Technometrics* 6(3):241– 252
- Embretson SE, Reise SP (2013) *Item Response Theory*. Psychology Press
- Erdogdu F, Erdogdu E (2015) The impact of access to ICT, student background and school/home environment on academic success of students in Turkey: An international comparative analysis. *Computers & Education* 82:26 – 49
- Ferguson R (2012) Learning analytics: drivers, developments and challenges. *International Journal of Technology Enhanced Learning* 4(5–6):304–317
- Ferguson R, Shum SB (2012) Social Learning Analytics: Five Approaches. In: *Proceedings of the Second International Conference on Learning Analytics & Knowledge*, ACM, pp 23–33
- Ferguson R, Macfadyen L, Clow D, Tynan B, Alexander S, Dawson S (2014) Setting Learning Analytics in Context: Overcoming the Barriers to Large-Scale Adoption. *Journal of Learning Analytics* 1(3):120–144
- Ferguson R, Clow D (2015) Examining engagement: analysing learner subpopulations in massive open online courses (MOOCs). In: *Proceedings of the Fifth International Conference on Learning Analytics & Knowledge*, ACM, pp 51–58
- Fonseca J, Valente MO, Conboy J (2011) Student characteristics and PISA science performance: Portugal in cross-national comparison. *Procedia-Social and Behavioral Sciences* 12:322–329
- Fournier H, Kop R, Sitlia H (2011) The Value of Learning Analytics to Networked Learning on a Personal Learning Environment. In: *Proceedings of the First International Conference on Learning Analytics & Knowledge*, pp 104–109
- French JJ, French A, Li WX (2015) The relationship among cultural dimensions, education expenditure, and PISA performance. *International Journal of Educational Development* 42:25–34
- Grawemeyer B, Mavrikis M, Holmes W, Gutierrez-Santos S, Wiedmann M, Rummel N (2016) Affecting Off-task Behaviour: How Affect-aware Feedback Can Improve Student Learning. In: *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge*, ACM, pp 104–113
- Gray G, McGuinness C, Owende P, Carthy A, (2014) A Review of Psychometric Data Analysis and Applications in Modelling of Academic Achievement in Tertiary Education. *Journal of Learning Analytics* 1(1):75–106

- Greks S (2009) Governing by numbers: The PISA effect in Europe. *Journal of Education Policy* 24(1):23–37
- Greller W, Drachsler H (2012) Translating Learning into Numbers: A Generic Framework for Learning Analytics. *Educational Technology & Society* 15(3):42–57
- Gupta D, Sharma A, Unny N, Manjunath G (2014) Graphical analysis and visualization of big data in business domains. In: *Big Data Analytics, Lecture Notes in Computer Science* (8883), Springer-Verlag, pp 53–56
- Hansen JD, Reich J (2015) Socioeconomic status and MOOC enrollment: enriching demographic information with external datasets. In: *Proceedings of the Fifth International Conference on Learning Analytics & Knowledge, ACM*, pp 59–63
- Hecking T, Chounta IA, Hoppe HU (2016) Investigating social and semantic user roles in MOOC discussion forums. In: *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge, ACM*, pp 198–207
- Hershkovitz A, Knight S, Dawson S, Jovanović J, Gašević D (2016) About” Learning” and” Analytics. *Journal of Learning Analytics* 3(2):1–5
- Hickey DT, Kelley TA, Shen X (2014) Small to Big Before Massive: Scaling up Participatory Learning Analytics. In: *Proceedings of the Fourth International Conference on Learning Analytics & Knowledge, ACM*, pp 93–97
- Hitlin S, Piliavin J (2004) Values: Reviving a dormant concept. *Annual Review of Sociology* 30:359–393
- Hofstede G (2011) Dimensionalizing cultures: The Hofstede model in context. *Online readings in psychology and culture* 2(1):8
- Hu X, Zhang Y, Chu SKW, Ke X (2016) Toward Personalizing an e-Quiz Bank for Primary School Students: An Exploration with Association Rule Mining and Clustering. In: *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge, ACM*, pp 25–29
- Joksimović S, Manataki A, Gašević D, Dawson S, Kovanović V, de Kereki IF (2016) Translating Network Position into Performance: Importance of Centrality in Different Network Configurations. In: *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge, ACM*, pp 314–323
- Kärkkäinen T, Saarela M (2015) Robust principal component analysis of data with missing values. In: *Lecture Notes in Artificial Intelligence* (9166), Springer International Publishing, pp 140–154
- Kjærnsli M, Lie S (2004) PISA and scientific literacy: similarities and differences between the nordic countries. *Scandinavian Journal of Educational Research* 48 (3): 271–286
- Kriegbaum K, Jansen M, Spinath B (2015) Motivation: A predictor of PISA’s mathematical competence beyond intelligence and prior test achievement. *Learning and Individual Differences* 43:140–148
- Kruskal W, Wallis W (1952) Use of ranks in one-criterion variance analysis. *Journal of the American statistical Association* 47(260):583–621
- Kumpulainen K, Lankinen T (2012) Striving for educational equity and excellence. In: *Miracle of education*, Springer, pp 69–81
- Laney D (2001) 3D data management: Controlling data volume, velocity and variety. Tech. rep., META Group
- Li N, Cohen WW, Koedinger KR (2013) Discovering student models with a clustering algorithm using problem content. In: *Proceedings of the 6th International Conference on Educational Data Mining*, pp 98–105
- Linnakylä P, Välijärvi J, Arffman I (2011) Finnish Basic Education – When Equity and Excellence Meet. In: *Equity and Excellence in Education: Towards maximal learning opportunities for all students*, Routledge, New York, pp 190–214
- Long P, Siemens G (2011) Penetrating the fog: Analytics in learning and education. *EDUCAUSE Review* 46(5):30–40

- López MI, Luna JM, Romero C, Ventura S (2012) Classification via clustering for predicting final marks based on student participation in forums. In: Proceedings of the 5th International Conference on Educational Data Mining, pp 148–151
- Marsman M (2014) Plausible Values in Statistical Inference. Universiteit Twente
- Merceron A, Blikstein P, Siemens G (2016) Learning Analytics: From Big Data to Meaningful Data. *Journal of Learning Analytics* 2(3):4–8
- Metsämuuronen J, Svedlin R, Ilic J (2012) Change in pupils' and students' attitudes toward school as a function of age - A Finnish perspective. *Journal of Educational and Developmental Psychology* 2(2):134–151
- Monseur C, Adams R (2008) Plausible Values: How to Deal with Their Limitations. *Journal of Applied Measurement* 10(3):320–334
- Morgan H (2014) Review of Research: The Education System in Finland: A Success Story Other Countries Can Emulate. *Childhood Education* 90(6):453–457
- Mukala P, Buijs J, Leemans M, van der Aalst W (2015) Learning analytics on Coursera event data: A process mining approach
- Musik A (2016) Philologenverband bezeichnet Pisa-Studie als Geldverschwendung. <http://www.deutschlandfunk.de/bildungsforschung-in-der-kritikphilologenverband.680.de.html?dram:articleid=347675>
- National Center for Education Statistics (2016) Program for International Student Assessment. <https://nces.ed.gov/surveys/pisa/>
- OECD (2011) Finland: Slow and Steady Reform for Consistently High Results. In: Successful Reformers in Education: Lessons from PISA for the United States, OECD, pp 117–135
- OECD (2012) PISA 2009 Technical Report. OECD Publishing
- OECD (2013a) PISA 2012 Results: Ready to Learn - Students' Engagement, Drive and Self-Beliefs (Volume III). PISA, OECD Publishing
- OECD (2013b) PISA 2012 Results: What Students Know and Can Do (Volume I) Student Performance in Mathematics, Reading and Science. v. 1, OECD Publishing
- OECD (2014a) PISA 2012 Results in Focus: What 15-year-olds know and what they can do with what they know. OECD Publishing. Paris: France
- OECD (2014b) PISA 2012 Technical Report
- Olsen RV (2005a) Achievement tests from an item perspective: An exploration of single item data from the PISA and TIMSS studies, and how such data can inform us about students' knowledge and thinking in science. PhD thesis, University of Oslo
- Olsen RV (2005b) An exploration of cluster structure in scientific literacy in PISA: Evidence for a Nordic dimension? *Nordic Studies in Science Education* 1(1):81–94
- Pardo A, Teasley S (2014) Learning Analytics Research, Theory and Practice: Widening the Discipline. *Journal of Learning Analytics* 1(3):4–6
- Peña-Ayala A (2014) Educational data mining: A survey and a data mining-based analysis of recent works. *Expert Systems with Applications* 41(4):1432 – 1462
- Picciano AG (2012) The Evolution of Big Data and Learning Analytics in American Higher Education. *Journal of Asynchronous Learning Networks* 16(3):9–20
- Piety PJ, Hickey DT, Bishop M (2014) Educational Data Sciences - Framing Emergent Practices for Analytics of Learning, Organizations, and Systems. In: Proceedings of the Fourth International Conference on Learning Analytics & Knowledge, ACM, pp 193–202
- Rasch G (1960) Studies in mathematical psychology: I. Probabilistic models for some intelligence and attainment tests.
- Rasmussen J, Bayer M (2014) Comparative study of teaching content in teacher education programmes in Canada, Denmark, Finland and Singapore. *Journal of Curriculum Studies* 46(6):798–818
- Rogers T (2015) Critical Realism and Learning Analytics Research: Epistemological Implications of an Ontological Foundation. In: Proceedings of the Fifth International Conference on Learning Analytics & Knowledge, ACM, pp 223–230

- Reich J, Tingley DH, Leder-Luis J, Roberts ME, Stewart B (2014) Computer-assisted reading and discovery for student generated text in massive open online courses. *Journal of Learning Analytics* 2(1):156–184
- Rutkowski L, Gonzalez E, Joncas M, von Davier M (2010) International Large-Scale Assessment Data Issues in Secondary Analysis and Reporting. *Educational Researcher* 39(2):142–151
- Saarela M, Kärkkäinen T (2014) Discovering Gender-Specific Knowledge from Finnish Basic Education using PISA Scale Indices. In: *Proceedings of the 7th International Conference on Educational Data Mining*, pp 60–68
- Saarela M, Kärkkäinen T (2015a) Analysing Student Performance using Sparse Data of Core Bachelor Courses. *Journal of Educational Data Mining* 7(1):3–32
- Saarela M, Kärkkäinen T (2015b) Do Country Stereotypes Exist in PISA? A Clustering Approach for Large, Sparse, and Weighted Data. In: *Proceedings of the 8th International Conference on Educational Data Mining (EDM 2015)*, pp 156–163
- Saarela M, Kärkkäinen T (2015c) Weighted Clustering of Sparse Educational Data. In: *Proceedings of the European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, pp 337–342
- Saarela M, Kärkkäinen T, Lahtonen T, Rossi T (2016a) Expert-based versus citation-based ranking of scholarly and scientific publication channels. *Journal of Informetrics* 10(3):693 – 718
- Saarela M, Yener B, Zaki MJ, Kärkkäinen T (2016b) Predicting Math Performance from Raw Large-Scale Educational Assessments Data: A Machine Learning Approach. In: *MLDEAS workshop of the 33rd International Conference on Machine Learning*, pp 1–8
- Sahlberg P (2011) *Finnish lessons*. Teachers College Press
- Santos JL, Klerkx J, Duval E, Gago D, Rodriguez L (2014) Success, activity and drop-outs in MOOCs an exploratory study on the UNED COMA courses. In: *Proceedings of the Fourth International Conference on Learning Analytics & Knowledge, ACM*, pp 98–102
- Schatz M, Popovic A, Dervin F (2016) From PISA to national branding: exploring Finnish education. *Discourse: Studies in the Cultural Politics of Education* pp 1–13
- Sedrakyan G, Weerd JD, Snoeck M (2016) Process-mining enabled feedback: Tell me what I did wrong vs. tell me how to do it right. *Computers in Human Behavior* 57:352 – 376
- Segedy JR, Kinnebrew JS, Biswas G (2015) Using Coherence Analysis to Characterize Self-Regulated Learning Behaviours in Open-Ended Learning Environments. *Journal of Learning Analytics* 2(1):13–48
- Siemens G (2013) Learning analytics: The emergence of a discipline. *American Behavioral Scientist* 57:1380–1400
- Siemens G (2014) The Journal of Learning Analytics: Supporting and Promoting Learning Analytics Research. *Journal of Learning Analytics* 1(1):3–5
- Siemens G, Baker RS (2012) Learning Analytics and Educational Data Mining: Towards Communication and Collaboration. In: *Proceedings of the 2nd International Conference on Learning Analytics & Knowledge, ACM*, pp 252–254
- Simola H (2005) The Finnish miracle of PISA: Historical and sociological remarks on teaching and teacher education. *Comparative Education* 41(4):455–470
- Skryabin M, Zhang J, Liu L, Zhang D (2015) How the ICT development level and usage influence student achievement in reading, mathematics, and science. *Computers & Education* 85:49–58
- Tømte C, Hatlevik O (2011) Gender-differences in Self-efficacy ICT related to various ICT-user profiles in Finland and Norway. How do self-efficacy, gender and ICT-user profiles relate to findings from PISA 2006. *Computers & Education* 57(1):1416 – 1424
- Trčka N, Pechenizkiy M, van der Aalst W (2010) *Process mining from educational data*. Chapman & Hall/CRC
- Väljjarvi J, Kupari P, Linnakylä P, Reinikainen P, Sulkunen S, Törnroos J, Arffman I (2007) *The Finnish success in PISA - and some reasons behind it: PISA 2003*. Jyväskylän yliopisto, Koulutuksen tutkimuslaitos

- Verbert K, Manouselis N, Drachsler H, Duval E (2012) Dataset-Driven Research to Support Learning and Knowledge Analytics. *Educational Technology & Society* 15(3):133–148
- Vogelsang T, Ruppertz L (2015) On the validity of peer grading and a cloud teaching assistant system. In: *Proceedings of the Fifth International Conference on Learning Analytics & Knowledge*, ACM, pp 41–50
- Waldow F, Takayama K, Sung YK (2014) Rethinking the pattern of external policy referencing: media discourses over the Asian Tigers: PISA success in Australia, Germany and South Korea. *Comparative Education* 50(3):302–321
- Wang Y, Paquette L, Baker R (2014) A longitudinal study on learner career advancement in MOOCs. *Journal of Learning Analytics* 1(3):203–206
- Wise AF, Shaffer DW (2015) Why Theory Matters More than Ever in the Age of Big Data. *Journal of Learning Analytics* 2(2):5–13
- Wise AF, Cui Y, Vytasek J (2016) Bringing order to chaos in MOOC discussion forums with content-related thread identification. In: *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge*, ACM, pp 188–197
- Worsley M, Blikstein P (2014) Analyzing Engineering Design through the Lens of Computation. *Journal of Learning Analytics* 1(2):151–186
- Wu M (2005) The role of plausible values in large-scale surveys. *Studies in Educational Evaluation* 31(2):114–128
- Wu M, Adams R (2002) Plausible values: Why they are important. In: *11th International Objective Measurement Workshop*, New Orleans
- Xing W, Wadholm B, Goggins S (2014) Learning analytics in CSCL with a focus on assessment: An exploratory study of activity theory-informed cluster analysis. In: *Proceedings of the Fourth International Conference on Learning Analytics & Knowledge*, ACM, pp 59–67
- Yates L (2013) Revisiting curriculum, the numbers game and the inequality problem. *Journal of Curriculum Studies* 45(1):39–51
- Ye C, Biswas G (2014) Early Prediction of Student Dropout and Performance in MOOCs using Higher Granularity Temporal Information. *Journal of Learning Analytics* 1(3):169–172
- Zaki MJ, Meira Jr W (2014) *Data Mining and Analysis: Fundamental Concepts and Algorithms*. Cambridge University Press
- Zhong ZJ (2011) From access to usage: The divide of self-reported digital skills among adolescents. *Computers & Education* 56(3):736 – 746