Open PISA: Dashboard for Large Educational Dataset

Avner Kantor
University of Haifa
Haifa, Israel
akantor@campus.haifa.ac.il

Sheizaf Rafaeli
University of Haifa
Haifa, Israel
sheizaf@rafaeli.net

ABSTRACT
International Large-Scale Assessments (ILSA) have a critical role in shaping education systems around the world. They impact local and national education policy and receive much attention in the media and the public discourse. However, the public has limited access to the results and cannot learn from them. Subsequently, the media might frame the results incorrectly. The transparency of ILSA is essential to the advancement of the public discourse. It requires easy access to data together with simple analysis tools. However, the complexity of ILSA makes it hard to understand and to analyze. Open PISA tries to deal with this challenge by developing a dashboard for the Program for International Student Assessment (PISA). It aims to guide users in the analysis of the dataset. This paper describes the dashboard design and insight based on collected users' responses. It hypothesizes that full transparency of the PISA dataset might be not achievable to the entire public. Further research is needed to evaluate how dataset analysis affects users' knowledge and opinions.

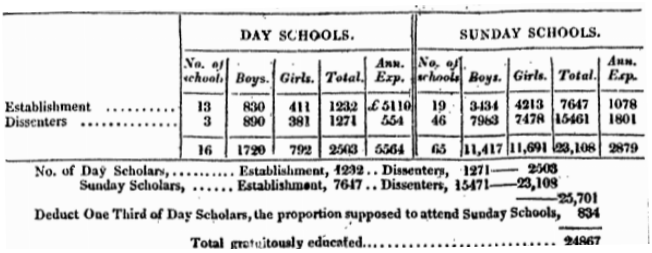
Author Keywords
Dashboards; citizen science; statistical literacy; PISA; international large-scale assessments; visualizations.

INTRODUCTION
The first edition of the Manchester Guardian in May 1821 published a four-page table of school sizes and the average annual expenditure in the Manchester area. It revealed to the public, although printed without comments, the educational inequalities in schools.

Two-hundred years later, the public is not exposed to datasets in general and educational datasets in particular. The exponential growth of collected data and the use of modern statistical methods make it hard to show the public raw data. Instead, the data is mediated to the public by reporters or scientists. Subsequently, the public discourse is controlled and framed by the media [15].

L@S '20, August 12–14, 2020, Virtual Event, USA

ACM ISBN 978-1-4503-7951-9/20/08...$15.00
https://doi.org/10.1145/3386527.3406721



	DAY SCHOOLS.					SUNDAY SCHOOLS.				
	No. of schools.	Boys.	Girls.	Total.	Ann. Exp.	No. of schools.	Boys.	Girls.	Total.	Ann. Exp.
Establishment	13	830	411	1232	£ 5110	19	3434	4213	7647	1078
Dissenters	3	890	381	1271	554	46	7983	7478	15461	1801
	16	1720	792	2503	5564	65	11,417	11,691	23,108	2879

No. of Day Scholars, ... Establishment, 1232 .. Dissenters, 1271 —— 2503
Sunday Scholars, ... Establishment, 7647 .. Dissenters, 15471 —— 23,108
—— 25,701
Deduct One Third of Day Scholars, the proportion supposed to attend Sunday Schools, 834
Total gratuitously educated..... 24867

Figure 1. Manchester Guardian, 1821.

In the last two decades, ILSA became central to educational policymaking around the world. The results of the tests are used to rank countries and compare their achievements. This leads to a major impact on the education systems that might be negative, as many argue in the Guardian [1]. At the same time, the public is exposed by the media only to part of ILSA results and furthermore cannot learn from them. Studies found that sometimes the media does not frame the ILSA results as expected [6].

Enabling the public easy access and simple learning of ILSA is essential to deliberative democracy and open science. The big data revolution of the last decade made this easier to apply. The technologies of data science, visualization, cloud storage and others change the way that data is accessible.

Open PISA (opisa.org) is a dashboard for the well-known program for international student assessment (PISA). It aims to enable learning at scale of the dataset using state of the art technologies. Policymakers, teachers, and school principals might use it to make better decisions based on evidence. The dashboard was designed to bridge the gap in knowledge and statistical literacy. This paper describes the dashboard design and insights based on users' responses.

PISA DATASET
PISA is a worldwide survey by the Organization for Economic Co-operation and Development (OECD). Its goal is to evaluate educational systems by testing skills and knowledge of 15-year-old students. It was first carried out in the year 2000 and then repeated every three years. In the 2018 cycle, over 600,000 students from 79 countries and economies participated.

The PISA dataset has an informative value at the national and international levels [7]. As Andreas Schleicher, Division Head of PISA, said it "... helped to change the balance of power in education by making public policy in the field of education more transparent and more efficient" [13]. Better

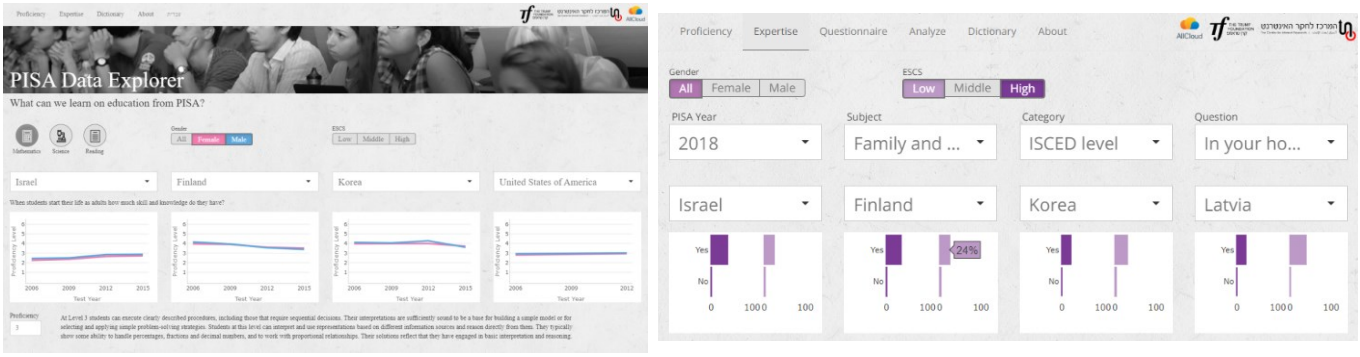


Figure 2. Two dashboard panels. Left: mathematics performance by gender; Right: distribution of study-desk at home by social and cultural status (ESCS).

understanding of the nature and effects of schooling are important for teachers, policymakers, and research communities. Data analysis is required for monitoring school reforms, predicting the effects of educational modification, and perspective on other countries' reforms [17].

A major criticism has been expressed concerning the way PISA results have been interpreted and used to introduce educational reforms [7]. Although the data is publicly available, its use is limited by available analytical tools for handling the complex sampling and testing procedures [2]. The dataset contains sampling variables and plausible values that cannot be handled by traditional statistics procedures. Also, the data is not straightforwardly accessible to users because it is organized in a large number of files in different categories [9] or with different variables [10].

DASHBOARD DESIGN

The PISA data is relevant among others to teachers, school principals, and parents. During the programming process the dashboard received responses from different types of users. It was distributed by the media, in academic seminars, and at conferences. The responses helped to define the expected users.

The dashboard developed using agile values as a feedback loop and adaptation cycle. Over the process users' responses helped to adapt the dashboard to their needs. This approach enabled by Shiny - an R package that makes it easy to build an interactive web app [3] combined with data analysis packages.

The dashboard includes several panels based on common PISA analysis: students' proficiency, students' expertise, students' and schools' principals' survey, and analysis of the relationship between survey variables and students' performances. Several hypotheses were made to support the learning of the panels.

Correct infographics. A full performance range has been used instead of showing only part of the scores. This way the differences in trends are less pronounced. Figure 2 demonstrates how the gender differences in mathematics in the chosen countries is negligible in this way. Secondly, the performance scores do not have a pedagogical meaning without comparison. Instead, proficiency levels were used together with a pedagogical explanation. Thirdly, the ranking of different countries is avoided because each country has its characteristics, and comparison might lead to the wrong conclusions. For the same reason, the average score of the performance of all countries is avoided.

Full data. The dashboard includes all the datasets and variables of PISA. The user can explore hundreds of variables from different cycles and the inquiry is not limited. An attached dictionary can be used to get the full list of variables. It was decided not to present the datasets from 2000-2003 due to a lack of statistical validity.

Inquiry methodology. A major method to understand the PISA dataset is by comparison. The design supports this type of inquiry. Figure 2 demonstrates a comparison of the students' performances in research cycles (only topics that were included in all research cycles were shown), by gender, by economic, social and cultural status (ESCS), and by countries (it was supported by showing several countries in the same line).

Statistics. The analyses are based on the PISA data analysis manual that "provides all the information required to understand the PISA databases and perform analyses" [4].

USERS' RESPONSE

Science-education graduates from several institutes in Israel experienced the dashboard. The students were instructed before the first use and throughout. Their responses raised several issues:

Infographics misinterpretation. Some of the students struggled to understand the infographics. Others preferred to use traditional PISA infographics (chart bars instead of trend lines).

Information overload. Enabling access to the full data made it hard to highlight important variables.

Inquiry methodology. The students struggled in understanding which questions the data might relate to and how to further research these.

Wrong conclusions. It was found that in many cases, the PISA naïve statistical method leads to the wrong

conclusions. For example, a linear analysis of the class size and mathematics performances shows a positive effect. That means larger classes have better performances.

Figure 3. Science-education graduates show the dashboard to faculty members at the University of Haifa.

DISCUSSION

The users' responses challenge the design hypotheses. Further research could clarify the reasons and how the design can be improved. Still, the responses provide useful insights for users' stratification and the differences between them.

Infographics misinterpretation might be a consequence of users' statistical literature or mistaken infographics. Education faculty members from several Israeli institutes connect this issue with the politics of visualization [16]. Each infographic means interpretation and the decisions of which axes, colors, countries and topics to use are all political. Ways to deal with that might be adding a disclaimer; visualization possibilities; giving users the ability to decide what to see; presenting all the variables and not only those that make statistical sense.

In the same way, there are politics of *inquiry methodology*. These could be improved by explaining the statistics concepts in use and explaining the PISA items [16]. It is hard to engage users with inquiries without self-interest [14]. Interactive infographics combined with questions, examples, and tutorials could encourage users to explore the data.

Information overload is a consequence of full access to the dataset variables. While researchers should have full data, it can disrupt other users' understanding. This can be solved by showing only data that can lead the user to interesting findings. This solution might lead to ignoring the survey panel in figure 2 because in most cases it does not include unique findings. Figure 4 shows a statistical data table with a guiding text. It is meant to help the users to compare interesting statistical values of variables. It required the designers to have a full analysis of the data before showing it to the public. But does that not miss the scientific idea of exploring the unknown? Also, who decides, what is considered to be interesting? A fact can be interesting to some, to others it can be insignificant.

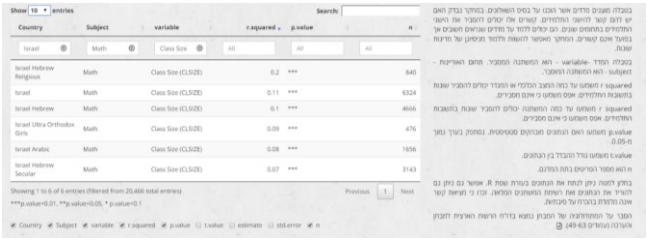


Figure 4. Statistical table to support variables comparison.

The real-life analysis of PISA and other ILSA datasets required different *statistical methods* for each question. Conventional interpretation can be misleading [12][5]. It can be another argument to PISA criticism [1]. Limited solutions could use an embedded development environment to enable straightforward access to the dataset or a crowdsourcing platform as Padlet to discuss the findings. These are limited to highly skilled users with programming and statistical knowledge.

CONCLUSION AND FURTHER RESEARCH

The dashboard, while it aims to enable public access to the PISA dataset, is in fact limited by the need for high skills and tremendous efforts [12]. While researchers don't need dashboard mediation, the public find it difficult to use. The important role of PISA and other ILSA datasets for policymaking requires finding ways to deal with this challenge although it may not be achieved.

Open PISA dashboard tested several attitudes to support the public as learners. This paper suggests several insights based on users' responses. Further research could clarify the reasons and how the dashboard can be improved. This would contribute to leverage data analyses and to increase human potential and interaction in the education discourse of the public.

This paper didn't discuss the effect of learning at scale on the public discourse. This is because of the challenge of reaching public engagement in investigation [11]. It is essential to study how learners' discoveries affect public trust in evidence-based education [8] and how it affects monitoring school reforms [17].

ACKNOWLEDGMENTS

Much appreciation to the people who provided helpful comments during the development of the dashboard design. The authors gratefully acknowledge the grant of the Trump Foundation for the dashboard development and scholarships, the support of the AllCloud company for cloud hosting, and the infographics advising of the Infoserviz company.

REFERENCES

[1] Paul Andrews, L Atkinson, S J Ball, M Barber, L Beckett, J Berardi, Y Zhao, D Berliner, E E Bloom, D Boudet, and others. 2014. OECD and PISA tests are damaging education worldwide--academics. *Guard.* 6, (2014). Retrieved from https://www.theguardian.com/education/2014/may/06/oecd-pisa-tests-damaging-education-academics

[2] Daniel H. Caro and Przemysaw Biecek. 2017. intsvy: An R Package for Analyzing International Large-Scale Assessment Data. *J. Stat. Softw.* 81, 7 (2017). DOI:https://doi.org/10.18637/jss.v081.i07

[3] Winston Chang, Joe Cheng, J Allaire, Yihui Xie, Jonathan McPherson, and others. 2017. Shiny: web application framework for R. *R Packag. version* 1, 5 (2017).

[4] Second Edition. 2009. *PISA Data Analysis Manual: SPSS, Second Edition*. DOI:https://doi.org/10.1787/9789264056275-en

[5] Yariv Feniger and Adam Lefstein. 2014. How not to reason with PISA data: an ironic investigation. *J. Educ. Policy* 29, 6 (2014), 845–855. DOI:https://doi.org/10.1080/02680939.2014.892156

[6] Mary Hamilton. 2017. Critical Studies in Education How International Large-Scale Skills Assessments engage with national actors : mobilising networks through policy , media and public knowledge How International Large-Scale Skills Assessments engage with national actors : mob. *Crit. Stud. Educ.* 58, 3 (2017), 280–294. DOI:https://doi.org/10.1080/17508487.2017.1330761

[7] Therese N. Hopfenbeck, Jenny Lenkeit, Yasmine El Masri, Kate Cantrell, Jeanne Ryan, and Jo Anne Baird. 2018. Lessons Learned from PISA: A Systematic Review of Peer-Reviewed Articles on the Programme for International Student Assessment. *Scand. J. Educ. Res.* 62, 3 (2018), 333–353. DOI:https://doi.org/10.1080/00313831.2016.1258726

[8] Shlomo Mizrahi and Yizhaq Minchuk. 2019. The relevance and reliability of performance information for accountability: A survey experiment exploring citizens' views. *Public Adm.* September 2018 (2019), 1–19. DOI:https://doi.org/10.1111/padm.12623

[9] Organization for Economic Co-operation and Development. Program for International Student Assessment, Data. Retrieved May 18, 2020 from https://www.oecd.org/pisa/data/

[10] Organization for Economic Co-operation and Development. Program for International Student Assessment, International Data Explorer. Retrieved May 18, 2020 from https://pisadataexplorer.oecd.org/ide/idepisa/

[11] Vineet Pandey, Justine Debelius, Embriette R. Hyde, Tomasz Kosciolek, Rob Knight, and Scott Klemmer. 2018. Docent: Transforming personal intuitions to scientific hypotheses through content learning and process training. *Proc. 5th Annu. ACM Conf. Learn. Scale, L S 2018* (2018). DOI:https://doi.org/10.1145/3231644.3231646

[12] Richard Rothstein and Martin Carnoy. 2014. "PISA Day"— An Ideological and Hyperventilated Exercise. *Economic Policy Institute*. Retrieved May 24, 2020 from https://www.epi.org/blog/pisa-day-ideological-hyperventilated-exercise/

[13] Andreas Schleicher. 2013. Big data and PISA. *HuffPost*. Retrieved July 22, 2013 from https://www.huffpost.com/entry/big-data-and-pisa_b_3633558

[14] Mike Sharples, Maria Aristeidou, Christothea Herodotou, Kevin McLeod, and Eileen Scanlon. 2019. Inquiry learning at scale: Pedagogy-informed design of a platform for citizen inquiry. *Proc. 6th 2019 ACM Conf. Learn. Scale, L@S 2019* (2019). DOI:https://doi.org/10.1145/3330430.3333642

[15] Miri Yemini and Noa Gordon. 2015. Media representations of national and international standardized testing in the Israeli education system. *Discourse Stud. Cult. Polit. Educ.* 6306, January (2015), 1–15. DOI:https://doi.org/10.1080/01596306.2015.1105786

[16] Michal Yerushalmi, Iris Wolf, Yael Kelly, Ayelet Baram-Tsabari, Iris Tabak, and Sheizaf Rafaeli. 2015. Visualizing and Presenting Educational Achievement. In *Doing Theory & Practice in LINKS together*. Retrieved from https://sites.google.com/a/edtech.haifa.ac.il/links-retreat-2015/visualizing-and-presenting-educational-achievement

[17] Larry D Yore, John O Anderson, and Mei Hung Chiu. 2010. Moving PISA results into the policy arena. *Perspect. Knowl. Transf. Futur. considerations Prep.* 8, 3 (2010), 593–609. DOI:https://doi.org/10.1007/s10763-010-9211-x