Comparative Causal Inference within PISA

PROPOSAL

Lauke Stoel (6899544)

Supervisors: Marieke van Onna (CITO) and
Remco Feskens (CITO and University of Twente)

Methodology and Statistics for the Behavioural, Biomedical and Social Sciences

Utrecht University

Date: 15/10/2021
Word count: 737

FETC-approved: 21-1939

Candidate journal: Large Scale Assessments in Education

1 Introduction

1.1 Background and research question

Myriads of valuable data on pupil's cognitive ability are periodically collected through International Large-Scale Assessments (ILSA's) such as the Programme for International Student Assessment (PISA). PISA is a triennial survey of 15-year-old students around the world that assesses their knowledge and skills in three core domains: reading, mathematics and science. Through PISA, the Organisation for Economic Co-operation and Development (OECD) aims to provide internationally comparable evidence on student performance (OECD, 2020). Ideally, these data could be used to compare different educational systems, to evaluate what features of each system lead to more favourable educational outcomes and inform national policy. However, it is impossible to conduct a randomised controlled trial and due to the observational nature the data, most methods used to analyse them are inept to distinguish descriptive quantities from causal effects (Rubin et al., 2004).

Rubin's potential outcome model provides a framework in which such causal questions could theoretically be addressed (Rubin et al., 2004). It entails a theoretical experiment where we treat the circumstance of interest one person is subject to as treatment condition A, and theorise that this person was simultaneously assigned treatment condition B, of which the data are non-existing. Those data are treated as missing and are estimated based on existing data from people with similar background characteristics, who are subject to treatment condition B. In a recent review of research on causal inference with large-scale assessments in education, Kaplan (2016) proposes an approach to causal inference within Rubin's potential outcomes framework that could theoretically expose causal links between educational outcomes and features of educational systems. However, his approach has a disclaimer: it has not yet been tested on existing data and currently available software is likely insufficient to fully execute the proposed statistical models (Kaplan, 2016).

The present thesis tests Kaplan's approach to causal inference on a currently relevant question within the Netherlands, namely whether it would aid the equality of educational opportunity (educational outcome) if the age at which pupils are selected into educational tracks is delayed (feature of an educational system). This topic is much discussed, as previous research suggests more heavily tracked educational systems are associated with a bigger negative influence of SES on the equality of educational opportunity (Bol et al., 2014; Veldhuis & Versteegh, 2021).

Kaplan's approach to causal inference could provide a more comprehensive way of addressing such questions, but is expected to have practical limitations. Another, theoretical, limitation is that Kaplan's approach is designed treat the ability level of a pupil as the outcome variable, whereas the above question requires an analysis with a relationship as the outcome variable, namely the one between SES and assessment outcomes. Therefore, the research question the current project addresses is: *to what extent can Kaplan's approach to causal inference be applied to inform educational policy regarding tracking, within the confines of currently available software?*

1.2 Analytical Plan

Data

We use the data from PISA 2018, which are publicly available. The data set contains the responses of 10,215 pupils from 79 participating countries and background information on the pupils, their parents and schools. We enrich this data set with relevant country-level characteristics, such as the age at which pupils are selected into tracks. We perform all analyses in RStudio (RStudio Team, 2020).

Applying Kaplan's approach:

Phase 1 - *Bayesian propensity score estimation (BPSE)* - risk: medium
Select relevant background characteristics to construct a propensity score equation for each pupil with, using R package "MatchIt" (Ho et al., 2011). For the initial analysis we use data from just two countries: the Netherlands and one other country, subject to educational tracking at a later age. Result: a subsample of pupils from both countries that is comparable on the selected background characteristics.

Phase 2 - *Estimating the causal effect* - risk: low
Identify the average treatment effect of being subject to tracking from a later age onward on ability estimates. Construct the Central Credible Interval around the estimate, indicating whether the effect is of significant influence.

Phase 3 - *Test against violations of assumptions* - risk: high
Perform sensitivity analysis on the treatment effect by including plausible hidden biases into the BPSE.

Extending Kaplan's approach:

Phase 4 - *Exploring limitations* - risk: high

a) Include multiple countries, exposing the challenge of incorporating multiple treatment conditions or treating tracking as a continuous variable.

b) Estimate the effect of tracking on the relationship between SES and educational outcome.

The end result will be a comprehensive overview of the strengths and limitations of Kaplan's approach as applied to real large-scale assessment data and its value in informing Dutch educational policy.

References

Alvarez, R. M., & Levin, I. (2021). Uncertain Neighbors: Bayesian Propensity Score Matching For Causal Inference [arXiv: 2105.02362]. *arXiv:2105.02362 [stat]*. Retrieved October 11, 2021, from http://arxiv.org/abs/2105.02362

Araujo, L., Saltelli, A., & Schnepf, S. V. (2017). Do PISA data justify PISA-based education policy? [Publisher: Emerald Publishing Limited]. *International Journal of Comparative Education and Development*, *19*(1), 20–34. https://doi.org/10.1108/IJCED-12-2016-0023

Bol, T., Witschge, J., Van de Werfhorst, H. G., & Dronkers, J. (2014). Curricular Tracking and Central Examinations: Counterbalancing the Impact of Social Background on Student Achievement in 36 Countries. *Social Forces*, *92*(4), 1545–1572. https://doi.org/10.1093/sf/sou003

Cordero, J. M., Cristóbal, V., & Santín, D. (2018). Causal Inference on Education Policies: A Survey of Empirical Studies Using Pisa, Timss and Pirls [_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/joes.12217]. *Journal of Economic Surveys*, *32*(3), 878–915. https://doi.org/10.1111/joes.12217

Frey, A., & Hartig, J. (2020). Methodological Challenges of International Student Assessment. In H. Harju-Luukkainen, N. McElvany, & J. Stang (Eds.), *Monitoring Student Achievement in the 21st Century* (pp. 39–49). Springer International Publishing. https://doi.org/10.1007/978-3-030-38969-7_4

Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2011). MatchIt: Nonparametric preprocessing for parametric causal inference. *Journal of Statistical Software*, *42*(8), 1–28. https://www.jstatsoft.org/v42/i08/

How, M.-L., & Hung, W. L. D. (2019). Harnessing Entropy via Predictive Analytics to Optimize Outcomes in the Pedagogical System: An Artificial Intelligence-Based Bayesian Networks Approach. *Education Sciences*, *9*(2), 158. https://doi.org/10.3390/educsci9020158

Kaplan, D. (2016). Causal inference with large-scale assessments in education from a Bayesian perspective: A review and synthesis. *Large-scale Assessments in Education*, *4*(1), 7. https://doi.org/10.1186/s40536-016-0022-6

Kaplan, D., & Kuger, S. (2016). The Methodology of PISA: Past, Present, and Future [Series Title: Methodology of Educational Measurement and Assessment]. In S. Kuger, E. Klieme, N. Jude, & D. Kaplan (Eds.), *Assessing Contexts of Learning* (pp. 53–73). Springer International Publishing. https://doi.org/10.1007/978-3-319-45357-6_3

Korthals, R. A., & Dronkers, J. (2016). Selection on performance and tracking. *Applied Economics*, *48*(30), 2836–2851. https://doi.org/10.1080/00036846.2015.1130789

Korthals, R. (2015). *Tracking students in secondary education : Consequences for student performance and inequality* (Doctoral Thesis) [ISBN: 9789053215364]. ROA. Maastricht. https://doi.org/10.26481/dis.20150618rk
Reeks in METIS: ROA Dissertation SeriesReeks nummer in METIS: 22

Kruijer, S., & van der Vegt, A. L. (n.d.). Kennisrotonde — Zorgt een brede brugklas voor een effectieve selectie? Retrieved October 11, 2021, from https://www.kennisrotonde.nl/vraag-en-antwoord/zorgt-brede-brugklas-effectieve-selectie

Martinková, P., Hladká, A., & Potužníková, E. (2020). Is academic tracking related to gains in learning competence? Using propensity score matching and differential item change functioning analysis for better understanding of tracking implications. *Learning and Instruction*, *66*, 101286. https://doi.org/10.1016/j.learninstruc.2019.101286

OECD. (2019). *PISA 2018 Assessment and Analytical Framework*. OECD. https://doi.org/10.1787/b25efab8-en

OECD. (2020). *PISA 2018 Results (Volume V)*. https://doi.org/https://doi.org/https://doi.org/10.1787/ca768d40-en

Pedemonte, F. B. (n.d.). Teacher professional development: A cross-national analysis of quality features associated with teaching practices and student achievement, 251
Interesting for cross-country analysis. Not for now.

Raudenbush, S. W. (2004). What Are Value-Added Models Estimating and What Does This Imply for Statistical Practice? *Journal of Educational and Behavioral Statistics*, *29*(1), 121–129. https://doi.org/10.3102/10769986029001121

RStudio Team. (2020). *Rstudio: Integrated development environment for r*. RStudio, PBC. Boston, MA. http://www.rstudio.com/

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, *63*(3), 581–592. https://doi.org/10.1093/biomet/63.3.581

Rubin, D. B., Stuart, E. A., & Zanutto, E. L. (2004). A Potential Outcomes View of Value-Added Assessment in Education [Publisher: American Educational Research Association]. *Journal of Educational and Behavioral Statistics*, *29*(1), 103–116. https://doi.org/10.3102/10769986029001103

Suk, Y., Kang, H., & Kim, J.-S. (2020). Random Forests Approach for Causal Inference with Clustered Observational Data [Publisher: Routledge _eprint: https://doi.org/10.1080/00273171.2020.1808437]. *Multivariate Behavioral Research*, *0*(0), 1–24. https://doi.org/10.1080/00273171.2020.1808437

van der Ploeg, S., & Dominguez Alvarez, L. (n.d.). Kennisrotonde — Draagt een later keuzemoment voor het voortgezet onderwijs, bijvoorbeeld op 14-jarige leeftijd, bij aan het schoolsucces van vooral leerlingen uit achterstandssituaties? Retrieved October 11, 2021, from https://www.kennisrotonde.nl/vraag-en-antwoord/latere-overgang-po-vo

Veldhuis, P., & Versteegh, K. (2021). Onderwijscrisis: Zijn brede brugklassen de oplossing? Retrieved October 4, 2021, from https://www.nrc.nl/nieuws/2021/04/16/onderwijscrisis-zijn-brede-brugklassen-de-oplossing-a4040183

Wu, M. (2009). A comparison of PISA and TIMSS 2003 achievement results in mathematics. *PROSPECTS*, *39*(1), 33–46. https://doi.org/10.1007/s11125-009-9109-y