Briefing Note: Update on methodology for creating a new socio-economic index for early learning

This report updates you on the development of a socio-economic index, or equity index, for early learning.

Summary

- The Sector Reference Group for the Review of equity and targeted funding in early learning expressed concern about applying the schooling equity index method in an early learning context. Their concerns are primarily related to the method requiring the use of an outcome variable, likely to be NCEA levels one and two.

- The Ministry responded by investigating other options for generating an early learning index. We tested the feasibility of the most promising option, Principal Component Analysis (PCA), by comparing a PCA-based schooling index with the established regression-based schooling equity index. The indices derived from the two methods are highly correlated.

- PCA would enable us to describe the socio-economic status of children attending an early learning service at an aggregated level, without requiring an outcome variable. The method is descriptive in nature, whereas the regression method is predictive. Both methods would utilise individual-level data from the Integrated Data Infrastructure (IDI).

- We presented the new method to the Sector Reference Group at the last meeting. Members were comfortable with the Ministry continuing to develop the early learning index using PCA.
Proactive release

**Agree** that this Briefing will be proactively released and any information which may need to be withheld will be done so in line with the provisions of the Official Information Act 1982

Agree / Disagree

John Brooker  
**Group Manager**  
**Education System Policy**

17/03/2021

Hon Chris Hipkins  
**Minister of Education**

25/3/21
Background

1. You agreed that targeted funding in ECE should be redesigned to allocate funding using individual level indicators of disadvantage in the Integrated Data Infrastructure (IDI) [METIS 1207062 refers]. This aligns with work to replace school decile funding with a new schooling Equity Index.

2. You also agreed that a Sector Reference Group comprising 15-20 members should be convened to support this work. Part of the Group’s role includes providing feedback and advice on different funding mechanisms and identifying potential risks and/or issues.

3. To date, we have held three meetings with the Sector Reference Group. At the second meeting we presented the methodology that was used to create the schooling Equity Index.

4. The schooling Equity Index uses regression, which is a predictive method. Using a historical population, a range of variables (such as parental qualifications and justice history) are linked to an outcome (NCEA achievement at levels one and two).

5. The model derived from the historical population is then applied to students currently enrolled in schooling to determine the level of disadvantage in each school. This will enable the Ministry to allocate resource to where it is most needed.

6. The Sector Reference Group had several concerns with the proposal to apply this methodology in an early learning context:
   - The Group was concerned about the outcome variable. They considered that by choosing an outcome variable, we would be making a judgement on what we value as success.
   - The Group felt that an educational achievement outcome such as NCEA would be inappropriate and inequitable, as it ignores other measures of success, such as cultural wellbeing. The Group felt this was particularly important in an early learning context, with representatives viewing child development as holistic in nature.
   - An alternative outcome variable that is closer to early learning would be the Strengths and Difficulties Questionnaire (SDQ) that is part of the Before School Check. The Group also considered this to be narrow and deficit focused. There are also technical issues with this variable, and it is not currently available for use in the IDI.

7. Given the issues raised by the Group, we investigated other methods for creating the early learning index that did not require an outcome variable.

An alternative method: Principal Component Analysis

PCA makes sense of complex data by identifying common themes

8. Our investigations led us to the Principal Component Analysis (PCA) method. This method is used to capture the collective features of many interrelated variables, without requiring an outcome variable.

9. PCA can be used to summarise multiple interrelated variables, producing a smaller number of “Principal Components.” One or more of these Principal Components can
then be used to describe the data. This process simplifies multi-dimensional data, whilst keeping as much of the detail as possible.

10. To look at a socioeconomic example, children of parents with higher incomes are likely to spend less of their lives on a benefit. Parents with high incomes are also likely to have higher qualifications and be less likely to have criminal convictions. Therefore, all these factors are correlated. PCA can be used to summarise these correlated variables into multiple Principal Components, which describe the underlying structure or essence of the data.

11. If a PCA method was used to create an Equity Index in early learning, the Principal Component/s would be used to describe the socio-economic status of the children enrolled with a service at an aggregated level. In effect, this would place all services on a continuum of the children’s socio-economic status. This differs slightly from the method used for schooling, as the schooling method creates a continuum based on how these socio-economic factors affect the children’s educational outcome.

**The two methods utilise individual-level data from the IDI, but PCA is descriptive whereas regression is predictive**

12. Compared to current systems, both regression and PCA methods would allow for a more nuanced approach to allocating funding and would be updated more regularly. Both methods would use individual-level data sourced from the IDI, capturing a large range of factors relating to disadvantage.

13. The key difference between regression and PCA is that one is predictive, whilst the other is descriptive. The regression method requires an outcome variable, whereas PCA does not.

14. The regression methodology, as used in the schooling Equity Index, relates socio-economic disadvantage to predicted educational achievement. PCA simply describes the pattern of socio-economic status for children in early learning without tying it to a child’s predicted future success in education.

**The two methods produce highly correlated indices when using schooling data**

15. The Ministry conducted a pilot study applying the PCA methodology to the existing variables that are used in the schooling Equity Index. This study enabled us to determine whether PCA would be technically feasible as a method for generating an early learning index.

16. The schooling Equity Index values produced using PCA were compared against schooling Equity Index values produced using the established schooling regression methodology. The study found there was a strong correlation between the outputs generated from the two methods. This indicates that the PCA-based index is meaningful and captures the socio-economic essence of the data.

17. There are also some differences between the two methods for the schooling data:

- PCA distinguishes between low and very low socio-economic status better than regression.
- PCA does not recognise some of the protective effects to the extent that the regression method does. For example, Asian students from low socio-economic backgrounds are statistically more likely to achieve NCEA than other students from low socio-economic backgrounds. The regression-based index explicitly accounts for this effect because all the variables are weighted based on their impact on
These protective factors are also included in the PCA method but may be weighted differently because they are not explicitly tied to an outcome. The PCA index therefore reflects more disadvantage for Asian children than the regression-based index.

**Presenting PCA to the Sector Reference Group**

18. We presented PCA as an alternative method for generating an early learning index at the most recent meeting of the Sector Reference Group on 18 February.

19. The Group was supportive of this method. Sector representatives considered that allocating funding based on the pattern of socio-economic status was appropriate in the early learning context. Whilst PCA does not use an outcome variable, life outcomes are likely to be correlated with socio-economic status to some degree. Therefore, funding based on socio-economic status alone is still likely to have positive effects for any given life outcome.

20. Existing equity-focused funding in early learning is generally used to mitigate the immediate effects of low socio-economic status, such as children arriving at the service without food. PCA would enable us to recognise services where there is a greater likelihood of this occurring, regardless of whether enrolled children are likely to go on to achieve NCEA. This was viewed as a positive feature of PCA.

**Next Steps**

21. We will continue to develop the early learning index, using the PCA method.

**Proactive Release**

22. We recommend that this Briefing is proactively released as per your expectation that information be released as soon as possible. Any information which may need to be withheld will be done so in line with the provisions of the Official Information Act 1982.