



**Cambridge  
Assessment**

# The use of longitudinal survey data in education research

*Conference Paper Abstract*

**Matthew Carroll**

Presented at the annual conference of the British Educational Research  
Association,  
University of Manchester, Manchester, UK,  
September 2019

**Author contact details:**

Matthew Carroll  
Assessment Research and Development,  
Research Division  
Cambridge Assessment  
The Triangle Building  
Shaftesbury Road  
Cambridge  
CB2 8EA  
UK

carroll.m@cambridgeassessment.org.uk

<http://www.cambridgeassessment.org.uk>

As a department of Cambridge University, Cambridge Assessment is respected and trusted worldwide, managing three world-class examination boards, and maintaining the highest standards in educational assessment and learning. We are a not-for-profit organisation.

**How to cite this publication:**

Carroll, M. (2019, September). *The use of longitudinal survey data in education research*. Paper presented at the annual conference of the British Educational Research Association, University of Manchester, Manchester, UK.

Theme: Research methodology in education

## **The use of longitudinal survey data in education research**

Important questions in education research concern the degree to which background and early experiences influence outcomes, and how young people's experiences of education change over time. To study these areas, datasets providing rich data across different ages are needed.

Surveys provide rich data, but in a single survey, data from different age groups can only be acquired by surveying different groups of participants, meaning that age is confounded by sample composition. Alternatively, administrative datasets provide repeated observations of individuals, but the data is limited in scope.

Longitudinal cohort studies survey participants repeatedly over time, providing rich data across multiple ages, without the confounding effects of changing sample composition (Cohen *et al.*, 2007). Whilst such studies are challenging and costly to run, they provide valuable data. However, it is not clear how resulting datasets are used within education research; understanding this could support effective data use and inform future research.

To explore the use of longitudinal datasets in education research, I carried out a literature review, examining publications using data from two recent, large-scale longitudinal cohort studies. The Millennium Cohort Study (<https://cls.ucl.ac.uk/cls-studies/millennium-cohort-study/>) follows children born in the UK between September 2000 and January 2002, and aims to study the influence of early life on later outcomes. Next Steps (formerly the Longitudinal Study of Young People in England; <https://cls.ucl.ac.uk/cls-studies/next-steps/>) followed participants aged 13-14 in 2004, up to the age of 25, and examined issues affecting young people and their transition to adulthood. I conducted searches on Google Scholar and Web of Knowledge between October 2017 and March 2018. I acquired sixty-one publications and working papers, reviewing each to identify the questions asked and methods employed. From this, I summarised the topics explored, common methodological approaches, and potential limitations. To the best of my knowledge, this is the first such review of this area.

I identified seven broad topics: "aspirations, expectations, and outcomes"; "correlates and drivers of attainment"; "ability grouping and perceptions of ability"; "subject choice"; "schools"; "disadvantage"; and "behaviour and personality". Some of these cannot be studied using administrative datasets (e.g., aspirations, personality). Others are frequently explored using administrative datasets (e.g., attainment, disadvantage), but the rich survey data allowed questions to be asked about variables that are otherwise unavailable (e.g., parental education). Therefore, a key use of longitudinal datasets is to access rich data that is simply not available elsewhere.

Most studies aimed to identify correlates of a focal outcome variable, utilising data from just one or two survey sweeps. Many used a two-stage methodology, calculating descriptive statistics then fitting regression models. Often, numerous background variables were included as controls, but their effects were not explored in detail. Whilst these studies

provide useful insights into large-scale associations, the frequent use of these methods means that some key aspects of the datasets are perhaps underused.

Some studies did use more complex aspects of the data. For example, structural equation models were used to explore interactions between variables, and the time-varying nature of the data was used to identify trajectories and transitions. Hence, with use of appropriate statistical methods, longitudinal datasets can be used to understand processes underlying large-scale associations.

Some limitations became apparent. As longitudinal studies are conducted over multiple years, by the time datasets are available and analyses are conducted, the education system can have changed, limiting the applicability of findings. Another limitation is that many studies rely on linked administrative datasets to provide examination results. However, as questions are raised about use of administrative datasets, it may be prudent to identify research questions that can be answered using survey data alone.

These findings indicate possible directions for future research. For effective policy or practical interventions, it may be necessary to understand what causes large-scale associations, so the rich survey data could be used to identify underlying mechanisms. Similarly, further use could be made of the time-varying aspect of the datasets; this is a unique feature, and the trajectories derived can be highly informative. Finally, it may be valuable to repeat some studies with data from more recent longitudinal surveys (e.g., Our Future; <https://www.ourfuturestudy.co.uk/>) to examine whether educational and economic changes have affected previously-identified associations.

In conclusion, this review highlights the value of longitudinal studies to education research. It shows how longitudinal survey data is used to study a broader range of questions than is possible with other data sources, but also that there is room to further utilise the complexity of the data.

## **References**

Cohen, L., Manion, L., & Morrison, K. (2007). *Research Methods in Education* (6th Edition). Abingdon, UK: Routledge.