## The bigger picture

Data mining tools and techniques were largely developed to serve life scientists who found themselves dealing with larger and larger datasets containing information about the sequences of DNA bases and protein amino acids. The size and complexity of such datasets demanded new tools, indeed a new field of bioinformatics, if researchers were to find answers to some of their existing questions and make the generation of new questions possible.

By bringing bioinformatics expertise to bear on education datasets, we have been able to exemplify how such findings can provide some insight for policy makers and practitioners. On its own, this insight has little utility unless it is used to shape and sharpen a shift in policy and practice. That is, the analysis becomes more useful when education professionals can use it to ask themselves new questions about what they want to achieve and how they might achieve it. For example, if the strategic intent of a policymaker, education leader or teacher were to increase student achievement in mathematics, this data may prompt them to ask questions about the way in which domain-specific anxiety is being addressed in classrooms. Should an educator then decide to intervene to lower levels of mathematics anxiety, the collection of new data becomes useful to indicate the extent to which the intervention has changed the levels of anxiety and met the strategic intent of increasing mathematics achievement. Well-established statistical techniques remain useful in analysing such data; the data mining techniques better serve the interrogation of large datasets, such as from PISA or from standardised tests, to identify issues and questions for educators to consider. The complexity of education means that in real-world situations, a reductionist approach that assumes simple relationships between cause and effect and clear correlations is unlikely to stand up to scrutiny or be useful to professionals’ everyday decision making. Data-mining helps to deal with this “messiness” to help educators direct their attention at the salient parts of large datasets that might otherwise be hidden below the surface.

In sharing the data-mining analysis with educators, it is important to help them to be vigilant for over-interpretation. It may be tempting for some to represent analyses such as ours as a “truth” rather than an indication of the importance of multiple dimensions within the data. When a particular, narrow analysis of data is used to push a singular interpretation over-reaching claims can sometimes result in the form of “the data says teachers should do this in their classrooms.” Of course, data will never be able to effectively dictate teachers’ practice. Even data specifically related to teacher practices, when analysed and interpreted into evidence, will have utility limited to the particular set of circumstances in which the data were collected. Space must be created for the findings to be processed through the professional judgement of educators and education leaders in order to develop an appropriate response in practice or policy.

Conversely, educators may be presented with evidence with too little scope for interpretation. For example, standardised tests may come with a demand that the school must do better next time, without providing any support for analysing the data further. When a breakdown of the data is provided it may be in terms of how the content of each question which, while it may have some use, it speaks only to a tactical response (e.g., spend more time on a particular topic). It does not speak to the existing strategic intent of the school or provide insight into what factors may need to become part of the future strategy for improvement.