



Automating data analysis of an ECD assessment tool

The [Early Learning Outcomes Measure \(ELOM\)](#) is a short psychometric child assessment tool commissioned by [Innovation Edge](#). Early Childhood Development (ECD) programmes can use this tool to assess the cohort of children in their programme to see whether they are developmentally on track for their age.

In total, 22 ECD organisations across South Africa use the ELOM tool to monitor children's performance at the ECD centres, playgroups, and home visiting environments that they support. Previously, the ELOM team used Excel to clean and analyse the data after each ECD programme submitted their assessment data, and then wrote short summary reports for each programme. This proved to be time-consuming and prone to errors. In 2018, Firdale Consulting was commissioned to create a solution to automate the reporting process, thus reducing the time burden on the team and improving accuracy.

How the ELOM works

The ELOM consists of two separate tests, one for children aged 50-59 months and one for 60-69 months. Children are assessed on their ability to perform a range of tasks related to five domains: gross motor development, fine motor coordination and visual motor integration, cognition and executive function, emergent numeracy and mathematics, and emergent literacy and language. The ELOM is app-based, allowing evaluators to easily upload the children's test scores for each domain onto tablets and store the data on a cloud-based server. After all children's tests are complete, the ECD programme can see the average ELOM scores for all children in the two age groups.





The ELOM is a powerful tool because of the use of age-appropriate benchmarks which the ELOM team has researched and verified. These benchmarks reveal how children should perform across the domains at each age, allowing ECD programmes to see if their class is, on average, achieving the standard, falling behind or at risk.

The need: Simple reports that highlight areas for improvement

Many of the staff at ECD programmes do not have a strong background in statistics, making complex statistical reports difficult to interpret. Instead, simple summary reports are required that show average performance per age group and highlight areas for improvement. Because each day of a child's life is critical to their development, it is important the programmes receive the reports as soon as possible after the assessments are complete so they can adjust where necessary.

The response: A report template that meets users' needs

Firdale Consulting was tasked with designing a report template. We reviewed the ELOM data, benchmarks and reports and met with the staff of an ECD programme in Khayelitsha that had just completed the ELOM test with their



children. Through an interactive workshop, we gained an understanding of what staff required from the reports and were able to gauge the average level of data literacy. Based on these insights, we designed a simple report that summarised the ELOM test results, highlighting any areas that required improvement. The prototype report was shared with the ECD programme staff and feedback was incorporated before the final design was decided.

The complication: Time-consuming and error-prone reporting processes

Creating these reports is a time-consuming process because it requires:

1. Cleaning and compiling raw ELOM score data,
2. Calculating statistics and producing graphs, and
3. Summarising these in a written report

Previously, the ELOM team relied on Excel to complete the first two steps. While Excel has the advantage of being easy to learn through its dropdown menu approach, its point-and-click interface means that for each report, you will need to "reinvent the wheel" when cleaning the data, calculating statistics and creating graphs.



In practice, this would mean that to create each ECD programme's report, we would need to import the raw ELOM score data, get rid of unnecessary characters through searching and replacing, convert columns stored as text into numbers, find and remove duplicate entries, and only then apply the formulae to produce the statistics and graphs we need.



Not only does redoing each of these steps add a lot of time to writing the reports, it also greatly increases the chance of mistakes and inconsistencies between the reports. In other words, Excel makes it a lot harder to ensure reproducibility.

The solution: Using a single statistical software package to automate cleaning and analysis

In order to reduce the amount of time the ELOM team spent producing each ECD programme's report, Firdale Consulting was hired to automate a significant proportion of the reporting process. We had two main options to overcome the issues facing Excel as a data analysis tool:

1. Write cleaning and analysis scripts in coding languages such as Stata, R, or Python
2. Use a visual data cleaning system such as OpenRefine or Wrangler

Compared to Excel, both options save time and provide better reproducibility, with the second option having the advantage of being more accessible to non-coders.



However, coding languages have the upper hand when it comes to managing inconsistencies in manually-entered data (e.g., "two" versus "2") and are better able to join multiple, large datasets together. Furthermore, many visual data cleaning systems do not offer data analysis, meaning we would still need to export the cleaned data into a programme like Excel to produce graphs and statistics.



For these reasons, we used Stata, a data analysis software that allows us to clean, merge, and analyse the data all within one platform. Although Stata has a steep learning curve, Firdale Consulting has strong in-house expertise in writing Stata code to clean raw data and produce insightful statistics and graphs. We also have expertise in R, which has the added benefits of being popular among data analysts and free to use.

We wrote Stata code to read the raw ELOM data (stored as Excel spreadsheets), perform numerous functions to clean the data, and then produce statistics tables (as Excel spreadsheets) and graphs (as PNG images). This automated output is added to the reports and forms the basis of the written content.



Using code to clean and analyse the data saves time because the same code can be used for every ECD programme's ELOM score data every year. Furthermore, because each report's statistics and graphs are produced following the exact same steps, consistency is ensured between reports and the chance of human error is eliminated. The analyses' accuracy is also improved because it is easier to trace errors when inspecting code than trying to remember the steps taken when using Excel.

The final product: An automated reporting system delivering easy-to-understand results with minimal time and effort

ELOM lets ECD programmes benchmark their children's performance against age appropriate national averages. After their children complete the ELOM assessment, ECD programmes want quick access to their results in a report that is easy to understand and highlights key areas for improvement.

The initial and critical task for Firdale Consulting was to engage with ECD programme staff to establish their level of data literacy and how best to present what they needed to know. The complexity of the ELOM data and the ECD programmes' need to access their reports within short timeframes made it necessary to use Stata code to partially automate the report generation.

At the completion of the project, Innovation Edge was equipped with a newly designed ELOM report and well-written Stata code that, with the click of a button, cleans data and produces the tables, graphs and other statistical output needed for the reports. This same code can be used again, thus shortening the delivery time of reports and providing the ECD programmes with simple yet meaningful measures of their children's development.