Identification of students at risk of early school leaving in Malta

Report for the Ministry of Education and Employment

This report was part of a project to support the Ministry of Education and Employment (MEDE) in complementing the existing data framework and developing a monitoring system for identification of students at risk of Early School Leaving (ESL) in Malta. ESL has been identified as a priority to be tackled, in line with the 2017 National Reform Programme of Malta, the 'Strategic Plan for the Prevention of Early School Leaving in Malta 2014' and the national EU2020 targets on reducing ESL.

The report provides proposals for a monitoring system to support early intervention in the case of ESL in Malta. The proposals are informed by analyses of administrative data for the cohort of students who should have completed the Secondary Education Certificate in the 2016/17 school year, identifying features that distinguished between those that did and did not obtain MQF Level 3.

It is the fifth and final deliverable for the project. It was preceded by three confidential reports (a review of existing data and data sharing procedures, business objectives for a proposed monitoring system), and a released report outlining some potential ESL interventions that might be introduced or extended in Malta.

The project was commissioned by the European Commission's Structural Reform Support Service (SRSS) and awarded to the International Association for the Evaluation of Educational Achievement (IEA).

The report was written by Dr Eemer Eivers, working on behalf of the IEA. It was submitted to MEDE in September 2019. The author acknowledges the considerable assistance of stakeholders, particularly the ESL Unit, in the development of the report. Any inaccuracies or misunderstandings contained therein are the responsibility of the author.

Table of contents

List of acronyms

ECEC	Early Childhood Education and Care
EMIS	Education Management Information System
ESL	Early School Leaving / Leavers
ESLU	Early School Leaving Unit
IMU	Information Management Unit
MEDE	Ministry for Education and Employment
MQF	Malta Qualifications Framework
NSO	National Statistics Office
NSSS	National School Support Services
SEC	Secondary Education Certificate
SIS	School Information Systems
QAD	Quality Assurance Department

Preamble

This document proposes a number of potential options for monitoring students at risk of Early School Leaving (ESL) in Malta. The first section introduces the current document and the broader project under which it falls. The second section outlines the methods used during this phase of the project, while the third outlines some general principles that should be considered when developing monitoring or early warning systems. The fourth section summarises the business objectives for a proposed new system (previously detailed in deliverable 3). Section five discusses the ESL risk indicators that should be considered for inclusion in a monitoring system. Finally, section six summarises the document and proposed options. Unless otherwise stated, Malta refers to the Maltese islands.

The document presented to MEDE included two additional sections, which outlined the preferred option (and likely costs) for a new monitoring system, and some alternatives to the preferred model. These sections are not presented here for reasons of commercial sensitivity.

1. Introduction

This report is the fifth and final deliverable in a project commissioned by the EU Structural Reform Support Service. The aim of this deliverable and of the wider project is to assist the Early School Leaving Unit (ESLU) in Malta's Ministry of Education and Employment (MEDE) in developing a centralised monitoring system for students at risk of ESL. The report is preceded by four earlier deliverables, three of which are not released publicly. These proposed a work plan for the project, summarised the current situation regarding the monitoring of ESL in Malta, and defined the business objectives for any monitoring system. The fourth deliverable and summarised some approaches to ESL prevention that may be effective in Malta – approaches either currently not in use in Malta, or in use, but not on a broad scale.

The project uses the European Commission's definition of early school leavers: those aged 18 to 24 with lower secondary education attainment at most, not currently in formal or non-formal education and training. In a Maltese context, this is taken as those who do not have at least five Secondary Education Certificate (SEC) passes at grades 1 to 7, or equivalent, and are not currently in any training or educational programme. The definition also includes early leaving from education and training. However, one of the main overall functions of this project is to facilitate *early* identification of students at risk of ESL. Thus, the primary targets of centralised monitoring of ESL Malta can be described as those not older than 16, where the focus is on work to prevent young people leaving school without any qualifications, or to direct them toward alternative educational pathways.

Deliverable

Deliverable 5 identifies possible scenarios for the development of a centralised monitoring system for students at risk of ESL, drawing on work conducted for earlier deliverables. This includes literature reviews, examples of good practices from other EU countries, comparison with other EU countries, expert consultation, identified business objectives for a new system, and analyses of where current practice and procedure meet or fail to meet the identified objectives.

2. Method

Four main tasks informed the identification of possible scenarios for a centralised monitoring system for students at risk of ESL in Malta. First, the contractor examined the literature in relation to ESL monitoring, with an emphasis on countries within the EU. Second, the contractor interviewed stakeholders in MEDE and other agencies to better understand their needs and their expectations from a monitoring system. Third, the contractor spoke to those implementing the monitoring of ESL in two other countries. This allowed for a practical counterpart to the research-based review from the broader literature. Rather than present the results of each of these activities separately, they form the basis for an integrated analysis of measures for a monitoring system in Malta.

In addition, attempts were made to identify potential ESL indicators, using data on students who had recently completed (or should have completed) compulsory schooling. The European Commission's definition of early school leavers relates to those aged 18 to 24, but, for the purposes of early identification and intervention, an earlier flag was needed. Thus, SEC 2016/17 outcomes were examined. Not attaining Malta Qualifications Framework (MQF) Level 3 was considered an earlier proxy for ESL status. Deliverable 2 had already identified some key sources of data relating to potential indicators, and this information was used to identify data sources from which a comprehensive student database might be compiled, including SEC outcomes.

3. General principles for monitoring systems

A centralised monitoring system is intended to be *predictive* of students' likelihood of dropout from school, based on certain characteristics (e.g., attendance history, behaviour, academic performance). There are many characteristics identified as predictors of ESL, but their relative predictive strength and the cut points for accurately identifying at-risk students can vary across contexts, even within the same country or state (e.g., Stuit et al., 2016). Thus, it is preferable to draw on local experience and data when developing early warning systems.

A monitoring system based on good local data can help to minimise *false positives* (incorrectly identifying students as at risk of ESL when they are not), while also correctly identifying the majority of those at high risk. For maximum efficacy, it should ideally be based on data from multiple cohorts (to counter any anomalous years) and include less overt forms of disengagement that may be predictive of ESL. The indicators used to predict ESL should also be checked against each new cohort of students, to gauge their continued relevance and accuracy in identifying students at risk. As there are also different "types" of ESL, it is also important to check that no sub-group is ignored by the monitoring system.

There are major challenges to implementing an early warning or monitoring system. First, to be effective, it needs to be based on data that are sound, recent and comprehensive, and such data are often not available. Second, sufficient resources must be provided to allow the system to be developed and rolled out across schools. Third, school and Ministry staff must be supported in using the system outputs and in providing appropriate responses to students flagged as at-risk.

A recent UNICEF (2016) report on frameworks for monitoring students at risk of dropout underscored the importance of a good quality Education Management Information System (EMIS) in supplying reliable data for planning, management, and decision-making. An EMIS should not just be a single national database, but a network of national, regional and school level systems. Rather than schools feeding data into a national level database, information should flow vertically (between local and national levels), and horizontally across all sectors. Further:

> "While database management, automation and Internet connectivity are important aspects of a modern EMIS, these will have little impact without considering also clear guidelines, training, understanding and acceptance by the people who collect, analyse and use the data for decision-making." (UNICEF, 2016, p. 10).

Such an EMIS does not currently exist in Malta. As noted in the second and third deliverables for this project, there is some vertical flow from schools to the Ministry, although outside of the state sector, little student-level data are shared. However, there is little downward vertical flow of data to schools, and very limited horizontal flow of data between agencies.

In essence, there is a silo-type approach to data, coupled with a lack of student-level data for a large minority of students. This was apparent in the difficulty the contractor had in establishing who the data holders were for various types of student data, and the subsequent difficulties in accessing any of that data, even in anonymised form.

NOTE: Many of the data access issues experienced by the contractor will be encountered if a monitoring system requires communication across agencies or bodies. Most bodies did not have a data-sharing culture. Many were nonetheless helpful and supportive, providing assistance once a justification for data-sharing was provided. On the other hand, a small number were less helpful.

Using the language of the World Bank's SABER (Systems Approach for Better Education Results) rubric for examining EMIS, Malta currently lacks an enabling environment for a centralised monitoring system, and limited use is made of existing data to aid decision-making. Specific to UNICEF's caution about the importance of human resource issues, current staffing levels and resources are not sufficient to support the introduction of a monitoring system. The capacity within MEDE and schools to analyse and use resultant data for decision-making would need to be significantly bolstered in order for a proposed system to be effective.

Developing a monitoring system without adequate support and without considering wider context will significantly impair its efficacy. Thus, UNICEF's (2016) eight-step general framework for developing a monitoring system for students at risk of ESL is of interest as it addresses the wider context in which a system would be located, and the prerequisites for an effective system. The framework is highly relevant to the current deliverable. The eight steps are listed next, followed by a more detailed description of what they involve:

1. Establish indicators, definitions and benchmarks.
2. Prevent, detect and resolve data inaccuracies.
3. Update EMIS to incorporate new indicators and methodologies.
4. Close gaps in vertical information flows between local and national level.
5. Close gaps in horizontal information flows through cross-sector collaboration.
6. Create an early warning / monitoring system.
7. Create automated reporting and analysis routines.
8. Develop and establish evidence-informed policies and interventions.

Establish indicators, resolve inaccuracies

As noted earlier, this project formally uses the European Commission's definition of early school leavers, which is based on those aged 18 to 24. While the definition facilitates cross-national comparisons, deliverable 3 highlighted how the definition caused practical problems for those working with students of compulsory school age in Malta. Stakeholders reported frustration with being unable to flag a student who had left school as ESL, because they were not in the 18-24 age range. Neither could they be described as NEET (not in education, employment or training) as they may be employed. Thus, the students could only be described as absent and subject to the standard procedures for monitoring absence. Relevant stakeholders felt that the capacity to apply another flag, such as *dropped out*, would be helpful.

In the context of an early warning and monitoring system, the Commission's definition might be replaced by one that focuses on those who should be in compulsory schooling but are not, or who do not plan to take the Secondary Education Certificate (SEC) examinations. This would also help address UNICEF's second step – address data inaccuracies – because it would mean that those who had left school were correctly flagged as such, rather than simply described as "absent".

Incorporate new indicators

To function as an ESL monitoring system, an EMIS needs data on all students who should be enrolled in school, including their attendance data and reasons for non-attendance. These data need to be at the level of the student as this facilitates the development of new indicators and methodologies. Student-level indicators are possible when disaggregated data entered into school content management systems are shared with the EMIS. In practical terms, a model such as MySchool or Klikks data feeding directly into a central monitoring system allows for development and real-time

monitoring of student-level indicators, whereas paper-based data or school-level aggregated data do not.

Thus, for optimum functioning, a monitoring system needs data-sharing links with school management systems. Data-sharing should be digital, record key student-level data, permit tracking of student from school to school, and be used in a consistent and reliable manner by schools. What constitutes key student data will vary depending on what local data have indicated are the most important student indicators. However, a general guide, they should include some basic student identifiers (unique ID, sex, date of birth, and contact details) as well as three core indicators often called the ABC of ESL (Academic achievement, Behaviour, Chronic absenteeism). The three indicators are often supplemented by D (for disability) and a number of other key variables such as family demographic information. Given the strong link in Malta between family socioeconomic status (SES) and educational outcomes, family information is likely to be a more important indicator in Malta than in many other countries.

Close gaps in vertical data flows

UNICEF identify a number of sources of data gaps that need to be addressed to support an effective monitoring system. Information which remains at the level of the school and is not shared vertically needs to be accessed. Thus, schools need to supply individual-level data on all students, all types of schools need to supply such data, and all key types of data need to be supplied. To maximise comprehensive and accurate data, data should be supplied electronically by schools directly to a central system.

Vertical information flows should be two-way. Those who supply data to a central system receive useful information in return – for example, being able to benchmark their school against national standards on selected indicators. However, data flows also need to address issues of data protection, especially regarding personal identifiers.

Close gaps in horizontal data flows

Horizontal gaps include a Ministry not being able to identify the entire school-age population (due to poor data sharing between government agencies) and poor cross-sector data sharing (leading to under-identification of risk). The former is not a significant issue in Malta, but the latter is. UNICEF recommends formal inter-agency data-sharing agreements that are cognisant of data protection sensitivities and data security, but which identify key data that needs to be shared and provide rationales for doing so. Data-sharing requires a common unique ID, which should (in theory at least) not be an issue in Malta as there is universal use of National IDs by government agencies. A number of different models can be used to share data, including permanent linking of databases, exported data files shared on a regular and scheduled basis, or an inter-agency database where different agencies contribute specified data to a pooled database.

Create an early warning / monitoring system

An early warning system needs to be able to facilitate an assessment of relative risk of ESL, and ideally, by the data provided, indicate the types of support an at-risk student might need.

Having established access to a pool of potential risk indicators, the first step in creating a monitoring system is to identify a small set of reliable risk indicators, and to use those to determine ESL risk. Those indicators can subsequently be applied to students to identify those most at risk of ESL. ESL risk can also be aggregated to the level of the school, to identify schools with a concentration of at-risk students. Such schools can add a "school context effect", meaning that attending these schools is in itself a factor in student risk and may be targeted for additional supports.

The nature and complexity of the variables used in a monitoring system is partly dictated by how it is accessed by schools. For example, a paper- or spreadsheet-based calculation of risk level would necessarily be quite basic, while whereas if risk level was automatically calculated in a database, it might permit inclusion of more indicators, weighted accordingly.

Subsequent to the identification of risk indicators, attention should turn to how to present risk information to school staff and other stakeholders. Ideally, student risk should be monitored throughout the school year, using real-time data. As such, stakeholders should ideally be able to access student risk reports from an EMIS (referring back to the importance of a two-way vertical flow of information). An effective monitoring system can generate reports at various levels (e.g., individual, school, college and national), using metrics such as overall risk scores and individual risk flags. For example, a student might be flagged as high overall risk of ESL, with specific flags for attendance and behaviour.

The number of indicators that feed into an early warning or monitoring system can vary considerably. With advanced statistical techniques and comprehensive and longitudinal datasets, it is possible to use predictive modelling to estimate individuals' probabilities of future outcomes such as becoming ESL. Multiple variables and iterative estimations can provide a *continuous* measure of risk (a probability between 0 and 1 of becoming ESL) for each student at given points in time. However, using such modelling on an ongoing basis requires significant statistical expertise and very detailed student datasets. In contrast, smaller numbers of indicators may only facilitate measures that are binary (at risk /not at risk) or categorical (e.g., low, medium or high risk). However, the data required are much easier to gather and system outputs easier to produce. Outputs also have the advantage of transparency, as it is easier for stakeholders to understand how students are flagged.

Where predictive modelling is not used, there is reasonable consensus that a small number of strong and reliable indicators is best (e.g., UNICEF, 2016; University of Chicago Consortium on Chicago School Research, 2014). Indicators selected need to be relevant, available (e.g., sensitive family information may not be accessible), and easily understood. There is some debate over the appropriateness of using non-malleable indicators (e.g., parental employment, migrant status) as schools cannot influence these characteristics. Nonetheless, such indicators are closely linked with risk of ESL – especially in Malta – and typically provide a better measure of risk. Frazelle and Nagel (2015) suggest that the choice of indicators be adapted over time, perhaps starting with the ABCD indicators and adding additional context indicators over time, using practical experience of their efficacy to decide if they should be retained.

Create reporting and analysis routines

Once a monitoring system is in place, outputs at various levels need to be developed. Reports can be targeted at the level of the student (identifying individuals who are at risk of ESL) and at sectoral and national level. For example, college principals might want to examine number of students currently flagged as at risk, by school, while MEDE might want to track changes over time in the percentage of students who left compulsory education without completing SEC examinations.

As well as reporting on overall risk of ESL, reports can target specific indicators of ESL (such as trends in absenteeism rates). With a two-way vertical flow of information, it would also be possible to examine the types of support provided to students flagged as at risk and, possibly, to draw some conclusions about the relative efficacy of supports and interventions.

It is important that system outputs are customisable and easy to use (e.g., interactive dashboard). However, it is equally important that supports are in place to ensure that outputs are regularly generated, that they are used to inform appropriate action and intervention.

Develop evidence-informed policies and interventions

The final element of a framework for an early warning or monitoring system is to use its outputs to inform policies and interventions. As UNICEF's framework document noted:

> "The practice of regularly and systematically using evidence for policy- and decision-making in ministries of education – or any other ministry – remains the exception rather than the rule. It requires a significant institutional and cultural shift." (2016, p 79)

The likelihood that the outputs of a monitoring system inform policy is dependent on a number of factors, including budgetary constraints and lack of inter-agency cooperation. However, providing clear reports and outputs, expressed in a manner that is understandable by policy-makers and practitioners, is key.

In contrast to policy, the implementation of interventions would typically be at the level of school and the individual student. Again, however, the extent to which the system outputs affect behaviour is dependent on the ease with which outputs can be generated and understood. Schools therefore need support in developing data analytics and data literacy, and using that information to implement individual and school-wide interventions.

4. Business objectives for a monitoring system in Malta

The business objectives for a monitoring system were described in detail in deliverable 3, and are therefore only summarised here. Based on extensive consultation with stakeholders, the key objectives for a monitoring system are for a **shared database** that pools information from multiple sources. This is expected to **reduce replication** of data and workload and to **minimise error** by drawing on existing verified data.

The database should provide **universal coverage** of students and include student- and aggregated school-level data. **Student data** should include information about the student's background, behaviour and academic performance. **School data** should include aggregated data on attendance, absence, achievement, and behaviour. This would not only allow identification of students at risk of ESL but also identify schools with clusters of at-risk students. The latter would facilitate MEDE in providing additional supports to such schools at a whole school level.

The system should provide different levels of **permissions-based access**, whereby those who have an agreed need to access specific data can easily get it in a usable format or presentation without the need for an intermediary. Permission rights should be clearly defined so that only those who require data on individual students can access it, with most users restricted to accessing aggregated data only. Teachers and school staff who provide data to the system should be able to have direct access to elements relevant to them. The system should be **easy to use**, and system outputs should support better use of existing data.

It should support the use of a **whole school approach** to tackling ESL by providing school and allied staff with outputs that are easy to access and relevant to local needs. For example, it should facilitate school staff in using school-level data to track issues such as attendance, performance and engagement, as well as assessment and examination outcomes. It should also support at least partial integration of common forms with the system, meaning that school staff should be able to draw on system data when completing forms. Relevant sections within MEDE (e.g., ESLU, Quality Assurance Department [QAD]) should be able to access **school-level reports** on attendance patterns, annual assessment data, and SEC outcomes.

The system should support the inclusion of **ESL risk indicators**, such as attendance, achievement, behaviour, and familial risk factors. For risk indicators to be effective, they must be based on data that are **consistent, accurate, and comparable**. This is not currently the case for some key indicators. Thus, to support a central monitoring system, MEDE may need to examine the lack of universal comparable measures of behaviour and achievement.

The system must provide high levels of **security** for any data held or shared and must be **GDPR compliant**. A well-developed integrated data system can provide better data security and privacy than disparate systems, which may require data to be shared in less secure formats such as emailed spreadsheet files. However, MEDE will need to develop business cases for certain data to be either collected and/or shared. It is likely that there will be resistance to collection or processing of some classes of data, resistance MEDE will need to overcome, in particular through proactive and transparent communication for the need and use of data.

In sum, the key business objectives are:

- Reduce replication
- Minimise workload
- Offer universal student data coverage

- Facilitate better use of available data
- Support whole school interventions
- Expand access to available data
- Integrate forms
- Use comparable indicators
- Support data protection and GDPR compliance
- Support data security.

In many regards, there is a large degree of similarity between the business objectives identified and the main steps in UNICEF's framework for an early warning or monitoring system. For example, the need for greater inter-agency communication and for schools to be able to access and use system outputs reflect the need to close gaps in vertical and horizontal information flows. Likewise, shared data as a means of minimising inaccuracies and replication is a common theme as is the need to protect the integrity and security of the data and the right to privacy of the data subjects.

> As part of the identification of a suitable monitoring system, it was intended to flag some risk indicators based on data from students in Malta. For accurate indicators, a comprehensive dataset for multiple, recent student cohorts was needed.
>
> The data that eventually became available (very late in the project timeline) only provided relatively comprehensive coverage for a subset of students. Even where information apparently had universal student-level coverage, it was not always reliable - e.g., prevalence of Special Educational Needs (SEN) was under-reported, and there were many missing values for familial SES.
>
> Also, data were provided for a single cohort only (those who were expected to complete the SEC in 2016/17). Basing analyses on this single cohort increases the risk of atypical factors influencing outcomes, and precludes the possibility of examining the effects of repeating a grade on completion rates.
>
> Bearing in mind these constraints, the section that outlines some of the variables that *might* be included as risk indicators has significant "health warnings" attached. The potential indicators and the attached warnings are common to all suggested models.

5. Preferred option

Confidential. Not released.

6. Other options for a monitoring system

Confidential. Not released.

7. Potential indicators of ESL Risk

The function of a monitoring system for students at risk of ESL is to identify risk. To do so, it must include variables that have not only been identified as associated with elevated risk of ESL, but are easy to quantify and to collect. For example, the number of books in a pre-schooler's home may be a good early risk indicator, but it is not the type of information that schools are likely to know or be able to collect. As such, variables of this type are not considered.

Based on work carried out for the second deliverable, a number of key sources of data that *should* be readily accessible and quantifiable were identified. These were:

- The E1 school content management system and its equivalents in non-state schools (containing attendance, behaviour and some achievement data).

- NSO data from the student census, SEC and SSC&P.

- Benchmark exam data held by EAU.

- Data on Scheme 9 uptake (an indicator of SES and income) held by NSSS.

- Parental employment data held by JobsPlus.

Based on discussions between the steering group, the contractor and NSO, it was decided to access the above data for a group of students to see what indicators of ESL risk might be identified. The target group of students was defined as those who would have been aged between 16 to 18 in 2016/17 school year (to allow examination of completion rates among the small number of students who repeated a grade). The exact data range was based on school age entrance criteria (from 1/1/2000 to 31/12/2001).

A complex procedure was used to ensure that student personal data were not compromised, and that no individual or school would be identifiable in the final dataset shared with the IEA. MEDE made a data access request to Identity Malta for a full list of National IDs for all those born in the specified data range. Identity Malta supplied a list to MEDE, who shared with the data holders. Data holders matched the IDs supplied to their data and extracted the relevant requested variables. They then sent the resultant data to NSO. Neither IEA nor MEDE received any data directly from the data holders, and never saw any data with an associated National ID. NSO acted as a central point for collating data from multiple sources. Finally, NSO anonymised the data and shared with the IEA only. MEDE have had no access to any data additional to that which was already held by MEDE.

Accessing data took *considerably* longer than anticipated, and the dataset was only received by the IEA on June 13, 2019. A review of the dataset revealed many shortcomings, some expected based on known data gaps, but others unanticipated. Due to delay in receipt of data, there was not sufficient time to address some gaps that might have otherwise been addressed. Any new data that could have been located and added to the dataset would have had to be submitted again for anonymity review by NSO, making it unlikely to be ready before the deadline (July 1, 2019) for the deliverable. Also, it would have required additional work from some data holders and from NSO. As the latter had already allocated a large amount of time and effort to the task (for which the contractor and steering committee are very grateful), a decision was made to proceed with the dataset available within the agreed project timeline. Thus, in the short time between receipt and reporting, only some basic analyses were carried out on the data, dealing as best possible with gaps.

Problems with data quality and coverage

There were a number of issues with the dataset received – some expected, and some not.

Expected issues

- Independent schools did not engage with the process, meaning the only data available for students in these schools are demographic and SEC outcomes.

- The absence of a universally administered exam (other than SEC) means a common achievement variable is not available.

- Many, but not all, Church schools granted access to Benchmark exam data. Thus, Benchmark data are available only for approximately 75% of targeted students.

- Scheme 9 data were unavailable for the cohort examined.

- Many, but not all, Church schools granted access to specified data from school content management systems. Due to variety in suppliers and the amount of work involved in requesting and collating data from multiple sources, only the two most popular were asked for data.

Unanticipated issues

- Identity Malta supplied National IDs for a single year only, rather than the two years requested. This means the effects of grade repetition cannot be examined. Also, the data are more exposed to being skewed by atypical factors associated with a single cohort.

- Poor National ID verification and matching. Almost 10% of National IDs supplied could not be matched to *any* existing data source.

  NOTE: Consultation with both NSO and MEDE suggest that these were data entry errors in the National ID database, rather than children who never engaged with state services. The unidentified IDs were removed from the dataset. However, there is a possibility that a small number of the IDs relate to real children, and that they were excluded from the final dataset.

- Under-representation of Third Country Nationals. Only a small number of the National IDs supplied were linked to Third Country Nationals, considerably fewer than expected. This may relate to the practice of assigning temporary "alien" National IDs to Third Country Nationals. However, it means that the experiences of Third Country Nationals are not adequately captured in subsequent analyses.

- Limited data from Church school content management systems. One of the two selected hosting companies did not supply any data. This reduces data coverage for Church schools. There may also be a slim possibility that schools using the sole responding system are somehow atypical of Church schools, generally – although this is unlikely.

- Problems with the reliability of data linking individual students to parental occupations, partly due to National ID linkage errors and partly due to the conflation of missing data with labour force inactivity (See explanatory box *ISCO and SES*).

- A small percentage (1.6%) of IDs that could not be given an ISCED attainment level (proxy for ESL) by NSO. Although likely to be ESLs, they cannot be treated as such, as their status in unknown. Thus, the analyses may slightly under-report ESL.

- It was only possible to establish change of school for those who had changed within the state sector or to/from the state to non-state sector. Movement within non-state schools could not be flagged.

Given the issues with the data, only limited analyses are possible, particularly in the case of those in Independent schools.

NOTE: Analyses that follow have a large caution attached, and should be taken as *possibly* representing the situation regarding ESL, but certainly not guaranteed to be accurate across all sectors.

ISCO and SES

International Standard Classification of Occupations (ISCO) codes are codes assigned to different types of occupations. The structure classifies occupations under 10 broad codes, which can be further broken down and linked to socioeconomic status. For example, ISCO Code 2 is the major group called "Professional", ISCO Code 22 is the sub-major group "Health Professional", code 222 is the minor group "Nursing" and 225 is the minor group "veterinarians". Normally, the more detailed three- or four-digit classification is used when transforming ISCO codes to a socioeconomic status scale.

However, data supplied for this project only included single digit ISCO codes. Broadly, lower codes are higher status occupations – e.g., code 1 is managerial and 2 is professional, whereas code 8 is for machine operators and code 9 is for elementary occupations such as cleaners and labourers, but there is a far from perfect correspondence between the single digit code and SES. Also, code 10 covers the armed forces, and can include very low status and very high-status occupations. In sum, with a single digit ISCO code, a lower code is usually but not always indicative of a higher status occupation.

As well as only receiving single-digit codes, there were a large number of students for whom no parental ISCO codes were supplied (e.g., close to half of maternal ISCO data were blank). This arose for two reasons. First, ISCO fields for students whose parents were inactive in the labour force (e.g., unemployed, parenting, retired, student) were left blank. Second, where parental and student National IDs could not be linked, ISCO fields were left blank. Thus, it was not possible to distinguish between unmatched students and those whose parents were inactive. For all of these reasons, relating student-level ISCO data to dropout risk was considered ill-advised.

However, SES is also known to be one of the key indicators of ESL, and ISCO codes were the only available source of relevant data. Thus, a school-level indicator was developed as follows. Each student for whom parental ISCO codes were available was assigned a "family" code, indicating if *any* parent was ISCO code 1 or 2 (i.e., usually the highest status codes). This provided a code for almost 85% of students, although missingness was noticeably lower among students in Church schools than in either state or Independent schools. Family codes were then aggregated to the level of school to establish the percentage of students in each school from an ISCO code 1 or 2 family background. The percentage was treated as a proxy value for high SES.

The procedure used is an unorthodox method of establishing the social class composition of a school, may over-state the percentage of high SES families in non-Church schools, and would not be recommended for future use. However, it was the only available option for the data provided.

Individual characteristics of "dropouts"

The term *dropout* is used to distinguish between the European Commission's formal definition of ESL (based on 18-24-year olds) and the proxy definition used here – students who were in Form 5 in 2016/17 and failed to achieve MQF level 3 in either the May or supplementary session. It does not refer to those who drop out from post-secondary institutions, although the term is often used in this context in Malta. Neither does it encompass students who repeated a year and were in Form 4 in 2016/17, because it was not possible to establish if they subsequently achieved MQF level 3. As they cannot be classified as either dropouts or completers, they are excluded from all further analyses in this section.

Region was closely linked with risk of dropout, with those in the Southern and Northern Harbour regions most likely to drop out (26% and 23%, respectively). Compared to a student from Gozo (7% dropout rate), a student in the Southern Harbour region was almost four times more likely to drop out (Table 3). Dropout was linked to **gender**, with rates higher among males than females (23% versus 14%, respectively).

Where available, Benchmark examination data were summed to provide a total **Benchmark examination score** (possible total of 300 marks). There were large differences on Benchmark scores obtained by those who subsequently did or did not achieve MQF level 3. Using a cut point of a total of fewer than 183 marks, almost half (49%) of students who fell below this cut point were dropouts. The cut point correctly identifies 86% of dropouts, and only 20% of completers. Unfortunately, we only have a single year of data for the Benchmark and it is unclear how comparable the marks assigned are from year to year. However, the score cut point is close to the bottom third of the score distribution (31.9% of all students obtained a total score below 183). Thus, a set value such as the lowest third could be applied across multiple years.

Looking at Benchmark examination results in each subject, the average gap between dropouts and completers was smallest for Maltese (about 19 marks), and highest for mathematics (almost 34 points), with a gap of 23 points for English. The finding that **mathematics** showed the largest points gap was true in both state and Church schools. Insufficient data were available to examine Benchmark performance for students in Independent schools.

Form 1 examination marks and Track information were combined to transform onto a scale of 0 – 3 for each of Maltese, English and Mathematics.[1] The scale was intended to be comparable to MQF levels (e.g., a value of 3 was assigned to a student taking Track 2 who had a score of at least 65%). Looking *only* at those who scored below the equivalent of Level 2 on English, 60% did not complete the SEC. The cut point correctly identified 84% of dropouts and 16% of completers. A similar cut point for mathematics correctly identified 90% of dropouts but also 25% of completers. For Maltese, only 68% of dropouts fell below Level 2, as did 11% of completers. Thus, performance on English seems to distinguish best.

Attendance data for the last year of primary school showed an average of 9 days absence for those students for whom data were available. Taking that as a cut point, 40% of completers and 77% of dropouts exceeded 9 days absence. For the first year of secondary school, a cut point of more than 20 days absences differentiated best between the groups (80% of completers and 39% of dropouts had absences below this cut). Attendance as an indicator is slightly problematic, however, as there

---

[1] Examination marks were mainly composed of common end-of-year examinations taken by those in state schools. However, they also include marks assigned by teachers in a number of Church schools. Although based on different examinations, the same broad relationship between marks and dropout held.

have been major reductions in absenteeism over the past few years (ESLU, 2018) and the data available for analyses here are heavily skewed towards state school students. Thus, a cut point based on partial data from a number of years ago may no longer be useful. A more realistic cut point might be the **20% of students with the poorest attendance,** nationally, at a given grade in any given year.

Dropouts had **consistently higher levels of absence** – from primary school to Form 5. The majority showed no major changes in rates of absenteeism, meaning that the absolute absence value is a better measure than *change* in absenteeism.

Perhaps due to how National IDs are initially assigned to those not born in Malta, there were relatively few EU and Third Country Nationals (TCN) identified in the database. Thus, while the number of **Maltese students** is sufficiently large to be able to say with reasonable confidence that the rate of dropout amongst them was approximately 18%, we cannot have confidence in the data reported for EU and TCN students. However, it seems likely that dropout rate is considerably higher in these groups than among Maltese students.

Information was supplied about some students' **SEN status**, but communication with relevant sections in MEDE indicate that SEN was under-reported. In the dataset supplied, just over 2% of students in Form 5 in 2016/17 were flagged as receiving some form of additional support. Of these, over half were dropouts. However, numbers involved are very small and it is possible that those flagged as SEN are atypical, meaning that it would be unwise to draw many conclusions about the relationship between SEN (more generally) and dropout from the data provided.

Table 3: Percentage of students who were classified as a "dropout", by various characteristics*

		% Dropout
Region	Southern Harbour	25.8
	Northern Harbour	22.7
	South Eastern	17.8
	Western	12.8
	Northern	15.6
	Gozo and Comino	6.8
Sex	Male	22.7
	Female	14.1
Exam data	Benchmark total below 183	49.3
	Form 1 English < MQF 2	60.0
Absence	**Absences reduced.** Use lowest 20%	
Nationality	Maltese	17.8
	EU (Under-reported, unreliable)	*35.6*
	TCN (Under-reported, unreliable)	*51.9*
SEN	*Under-reported, unreliable*	*59.3*
Changed school	*Only between or in/out of state schools*	*33.8*
Repeater	**Cannot be established**	–
SES	**Data not reliable**	–
Behaviour rating	**Data not comparable**	–
Behaviour supports	**Data not available**	–

*Subject to aforementioned cautions about quality and comprehensiveness of the dataset.

As noted earlier, data were supplied only for a single cohort of students (those who should have completed the SEC in 2016/17). As SEC outcomes were therefore unknown for students in this group who had **repeated a year**, it was not possible to establish their dropout rate. However, research from other countries suggest grade repetition is associated with elevated risk of ESL.

Information about student or **family SES** was also unreliable, as it was not possible to distinguish between students whose parents were not active in the labour force and students for whom data entry errors meant their National ID could not be linked to that of their parents. Thus, for example, maternal occupational ISCO codes were absent for 44% of students, but it was not possible to establish which were inactive and which were missing data. However, looking *only* at those students for whom ISCO codes were available, dropout was highest amongst those whose parents were engaged in manual labour types of work (e.g., factory workers, machine operators, and elementary occupations such as cleaners and labourers). It was lowest amongst students whose parents were professionals.

Limited information was available about students who **changed school** during their secondary school years. It was possible to identify those who had moved between different *state* schools, and those who had moved into or out of the state sector. However, no information was available for those who might have moved within the non-state sector. Excluding those who transferred to the Alternative Learning Programme, the approximately 4% who either moved within/in/out of the state sector had well above average rates of dropout. As numbers were relatively small, all groups were combined. Overall, 34% of those who changed school (either from a non-state to a state school, moving between state schools, or moving out of the state school sector) were classified as dropouts.

Information about student **behaviour** was available from the SSC&P (a 3-point scale) and from schools' own positive and negative behaviour ratings used in the Klikks system. The SSC&P scale was highly skewed and too truncated to be useful, as very few students in Form 1 and 2 (less than 5%) were rated as "Needs to improve". In contrast, the positive and negative behaviour scores used in Klikks varied so much between schools that combining them did not provide useful data. Thus, while poor behaviour is highly likely to be an indicator of dropout, the available dataset does not contain data of sufficient quality to provide a reliable cut point. Likewise, student-level data on use of nurture groups or learning zones was unavailable.

Sector and school characteristics

As noted, there were significant gaps in the dataset supplied to the contractor, particularly for non-state schools. This makes it difficult to draw any firm conclusions about school and sector characteristics related to risk of ESL. Nonetheless, one conclusion that can be drawn is that there are large differences in rates of ESL by school sector, even after ALP data are removed from the state sector. There is also variation between individual schools in rates of ESL, both within and across sector. However, it is also clear that much of these differences can be linked back to pre-existing differences in intake. One of the clearest differences between those who did or did not drop out is performance on the Benchmark examination. There are very large differences between secondary schools, by sector, in average Benchmark scores obtained by incoming Form 1 students. In other words, it may not be the case that different secondary school sectors lead to different student outcomes, but rather that different types of students attend different school sectors.

With the partial data available for this project, disentangling the "value added" by schools from the pre-existing differences in intake characteristics was not feasible in the limited time available. Bearing this in mind, the remainder of this section outlines some sectoral differences for student SES, Benchmark outcomes, attendance, and completion rates.

As noted, there were issues with attaching parental ISCO codes to students, given the lack of distinction between those for whom parental data were missing and those whose parents were not active in the labour force. Also, Scheme 9 data are unavailable for the cohort in question. However, SES is an important correlate of ESL, so the data were re-examined to see if *any* reliable measure of SES could be identified. A school-level aggregate was created using ISCO data, based on the assumption that the problem with non-linked student-parent IDs was random (i.e., due to data entry errors, rather than systematically biased by social class).

Student family ISCO codes were aggregated to the level of school to establish the percentage of students in each school whose parents had been assigned a code 1 or 2 – i.e., were from a managerial or professional family background. The percentage is treated as a proxy for higher SES. As noted earlier, this was an unorthodox method of establishing the social class composition of a school, borne out of lack of other options. It is not a recommended approach and *considerable* caution needs to be used when interpreting any SES-related data.

Across all sectors, approximately 15% of students could not be assigned any family ISCO code. Excluding these students, approximately one-third (32%) of the remainder were from families with at least one parent assigned ISCO code 1 or 2. However, the extent of missing data varied considerably by sector, from just over 6% in Church schools to just over 20% in other sectors. A possible explanation is that almost all Church school students were Maltese, and neither they nor their parents would ever have had temporary National ID numbers (which complicate the matching process). Whatever the reason, the differences in missing data means that while the estimated percentage of **high SES families** in Church schools is likely to be accurate, the estimated percentages in state and Independent schools are less so. On the balance of probability, the rates may be slightly over-stated. With this caveat in mind, of students whose family ISCO status was known, 20% of students in state schools were in the higher SES category, compared to 39% in Church and 73% in Independent schools (Table 4).

The end of primary school **Benchmark examination** and data on **absences from Year 6** both provide information about pre-existing differences in the characteristics of the student body, by secondary school sector. As with SES, there were large differences in the benchmark scores for students who went on to attend state versus Church schools (an average of 193 marks in total, versus 220, respectively). Benchmark data were also available for some students who subsequently attended Independent schools, but the numbers involved are too few to be able to draw any reliable conclusions. Regarding **absence**, those who transferred to state schools missed an average of almost 14 days in the final year of primary school, compared to 7.5 days absences for those who transferred to Church schools.

Given the large differences in the nature of their respective intakes, it is not surprising that there are also **large differences by sector in the percentages of students who did not attain MQF Level 3** (i.e., "dropouts"). In state schools, 24% of students were dropouts, compared to 6% in Church schools and just over 7% in Independent schools.

> The percentages shown in Table 4 are based on the schools in which students were enrolled in their *final year* of secondary school. Anecdotally, the contractor has been told that in any given year, a small number of students at risk of dropout transfer from non-state to state schools. If this is the case it means that the percentage dropout shown for state schools is marginally inflated by these transfers, whereas the percentages for Church and Independent are marginally depressed.

Across all sectors, there was a negative relationship between school-level dropout rates and school-level SES. For Church and state schools (reliable data unavailable for Independent schools) there were also negative relationships between school-level average benchmark results and school-level absences at primary and secondary. In other words, fewer students dropped out of schools with high average SES, above average Benchmark results and good attendance levels.

Table 4: Summary characteristics of secondary schools, by sector, excluding ALP

	State	Church	Independent
*High SES students**	*20%*	*39%*	*73%*
Mean Benchmark total	193 marks	220 marks	*Insufficient data*
Mean Year 6 absences	13.7 days	7.5 days	*Insufficient data*
% "dropout"	24.2%	5.9%	7.4%

*Caution needed due to large missing data. May over-state percent of high SES in non-Church schools.

Interaction of school & individual characteristics

To illustrate how individual and school characteristics interact, two examples are provided.

- On average, students from the Harbour areas had a dropout rate of 24% (including students attending the ALP). However, the rate of dropout is related to the average SES of the school attended. If schools are split into those with low, medium, and high SES intake, there are large differences in rate of dropout for Harbour area students. In high SES schools the rate was only 7%, compared to 19% in medium SES schools, and 38% in low SES schools.

- The large overall differences in rates of dropout by sector are considerably reduced when only schools with *an at least average* percentage of high SES families are considered. For this subset of schools, dropout from state schools is 10%, compared to 5% in Church and 4% in Independent schools

Summary of potential indicators

Based on the data provided, reliable measures of some key indicators are missing – e.g., nationality, SES, and any *early* measures of attendance or achievement. However, based on what has been gleaned from the incomplete data file, **student-level flags** to consider are:

- Region (both Harbour regions).
- Attendance (an approximate value of the lowest quintile, nationally, at that grade).
- Achievement (an approximate value of the bottom third nationally, at that grade).
- Gender (males have elevated risk).

Based on evidence from broader research, student behaviour, lateness, SES, SEN status and nationality are all also likely to be indicators of ESL risk.

Other student-level required data are student and parental National IDs, linked to the Corporate Data Repository to minimise data entry errors. The inclusion of parental National ID can also facilitate measurement of SES.

School-level flags to consider are:

- Low average school-level SES (assessed using Scheme 9 data, or data obtained using parental National IDs).
- Average Benchmark performance in the bottom third of distribution.
- High school-level ESL rates (secondary schools only).
- High school-level absence rates.

The number of indicators used, and any relative weightings assigned, is dependent on the methods planned for monitoring. For example, it is unlikely that a complex measure with multiple, differently weighted variables will be used correctly if the monitoring system consists of manual data entry and calculations by school staff. As a rule of thumb, the more basic the monitoring system, the simpler the indicators should be.

It is important that indicators are as comparable as possible – particularly for attendance, achievement and behaviour. At the most basic, student-level data must include achievement scores and a total number of absences, all from as early as possible in a student's school life. It should also include agreed measures of lateness and behaviour. The behaviour rating scale currently used for the SSC&P is too truncated to provide adequate discrimination, whereas scales used in Church schools are too varied to be useful. That said, it should be possible to agree a common behavioural rating scale that incorporates both positive and negative measures.

However, for a monitoring system to be effective, it must also take school context into account.

8. Summary

This document is the fifth and final deliverable in a project to assist the development of a centralised monitoring system for students at risk of ESL in Malta. It identifies a number of possible scenarios most suitable in a Maltese context, drawing on earlier work which examined existing data relevant to ESL in Malta, identified the business objectives for a monitoring system, and suggested some approaches to tackling ESL that might work in Malta.

Although grounded in the current situation in Malta, the report draws heavily on two external sources in evaluating what would be most appropriate and effective, and what is required to implement a monitoring system. The external sources are the World Bank's rubric for evaluating education management information systems (Hamid, 2014) and UNICEF's framework for monitoring educational participation (UNICEF, 2016).

These two sources, in conjunction with the business objectives for a new system as identified by key stakeholders in Malta, informed the identification of a preferred approach. Certain common themes emerged, including the need for student-level data, for greater inter-agency communication, and for schools to *use* system outputs. Information should flow both vertically and horizontally, to and from a central system. Securely shared data as a means of minimising inaccuracies, reducing workload and replication was a common theme, as was the need to balance sensitivities about the right to privacy of the data subjects against the need for a Ministry to access data in order to fulfil its basic functions.

Identifying indicators of risk

Part of the final deliverable included accessing and analysing data for a cohort of students in Malta, to identify some of the indicators of ESL that might be included in a monitoring system. Accessing the data was a long and complex process. Although initiated in October 2018, the data were only received in mid-June 2019. The received dataset had a number of gaps and issues that meant the development of a complex predictor model was not advisable. For example, no data were available for repeater students, and reliable data were not available for student SEN status, nationality, or behaviour.

However, a number of risk factors *were* identified, including poor academic performance on the Benchmark examination, residing in one of the Harbour regions, poor attendance, and attending schools with a high average rate of ESL or low average SES. Rate of "dropout" (the proxy indicator used for ESL[2]) was higher among males than females, and among those who attended state schools. The data also showed differences in the intake characteristics of secondary schools, by sector, including sizeable differences in average primary school attendance and achievement, and family SES. Thus, while rates of ESL differ by sector, much of the differences *could* be attributable to pre-existing differences in student characteristics.

The difficulties in accessing data flagged a number of issues that need to be addressed if a monitoring system is to be effectively implemented. Most importantly, MEDE need to develop a legal framework that supports the collection and sharing of data that will be required for *any* monitoring system for ESL. The framework must legislate for the collection and dissemination of data by specified agencies for specified purposes. It must mandate schools to supply specified data, identify the need for vertical and horizontal information flows, address GDPR and privacy concerns,

---

[2] The formal definition of ESL covers those aged 18-24. However, an early warning system is necessarily directed at a younger age group. Thus, for the purposes of this project, those who did not successfully complete the SEC in Form 5 were classified as "dropouts", which was treated as a proxy for ESL.

and designate responsibility for data collection, analysis and dissemination. It will also need to address the current silo-type approach to data within some state agencies, and the lack of student-level data for a large minority of students. Without student-level data, student-level monitoring is not possible.

In addition, the report recommends that MEDE liaise with leaders in non-state sectors *before* rolling out any monitoring system, and that the benefits to schools of supporting such a system are clearly flagged. As well as reducing rates of ESL in all schools, benefits might include using it as a basis for access to additional resources, and perhaps some core resources. This may counter perceptions that sharing data is a reduction in schools' autonomy.

Finally, the report recommends that sufficient staffing is put in place to adequately support a monitoring system. This is likely to include not only technical support staff to assist schools during roll out and initial use of the system, but also a project manager, a statistician and a dissemination officer. The latter two roles will prove particularly helpful in developing the capacity of MEDE and schools to mine the resultant data and identify school- and system-level issues, and to help schools use the data to inform whole-school approaches to tackling ESL.

Proposed option

Due to commercial sensitivities, details of the proposed option have been removed from this report. The remainder of this section outlines the proposed option, but only in very broad terms.

The option recommended is to have a single school management system covering all schools in Malta, with a number of high-level dashboards, some extraction functions (for common forms), and a small number of specified ESL indicators as required fields. The system should be supplemented by two non-integrated additions– the facility to import data from Early Childhood Education and Care (ECEC) settings and export/import data for SEC registrations.

The proposed option meets the business objectives identified by key stakeholders. It would be a single database:

- with relevant and centrally accessible data, available to MEDE and staff in related agencies, facilitating vertical and horizontal information flows.

- drawing on real-time student- and school-level data to monitor student risk, while also having access to students' historic data.

- linked directly with school information and content management systems.

- containing data on a number of indicators of risk of ESL.

The main strength of the preferred option is that it would provide in-depth student-level data for all students, and aggregated school-level data for all schools. It would reduce replication of work and of data, minimise opportunity for error, provide real-time data, and be a significantly more secure way of sharing data than current practice.

The main obstacles identified are likely stakeholder resistance and the lack of a legal framework underpinning the collection of such data by MEDE. These obstacles, however, are likely to apply to any model for a monitoring system that accesses student-level data from schools and possibly from other agencies.

References

Dubay, L., & Holla, N. (2016). *Does attendance in early education predict attendance in elementary school? An analysis of DCPS's early education program*. Washington, DC: Urban Institute.

Early School Leaving Unit (2018). *A report on the implementation of the strategic plan for the prevention of early school leaving in Malta 2014*. Malta, MEDE.

European Commission (2019). *2nd survey of schools: ICT in education. Objective 1: Benchmark progress in ICT in school*. Luxembourg: European Commission. doi: 10.2759/23401

Ferguson, R., Brasher, A., Clow, D., Cooper, A., Hillaire, G., Mittelmeier, J., Rienties, B., Ullmann, T., Vuorikari, R. (2016). *Research evidence on the use of learning analytics - implications for education policy*. R. Vuorikari, J. Castaño Muñoz (Eds.). Joint Research Centre Science for Policy Report; EUR 28294 EN; doi:10.2791/955210.

Frazelle, S. & Nagel, A. (2015). *A practitioner's guide to implementing early warning systems (REL 2015–056*). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Northwest. Retrieved from http://ies.ed.gov/ncee/edlabs.

Hamid, H.A. (2014). *SABER: What matters for most education management information systems - a framework paper (English)*. SABER working paper series no. 7. Washington, DC: World Bank Group.

Siemens (2011). *Learning and Academic Analytics.* Blogpost 5 August 2011 https://www.learninganalytics.net/uncategorized/learning-and-academic-analytics/

Stuit, D., O'Cummings, M., Norbury, H., Heppen, J., Dhillon, S., Lindsay, J., & Zhu, B. (2016). *Identifying early warning indicators in three Ohio school districts (REL 2016–118).* Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Midwest. http://ies.ed.gov/ncee/edlabs/projects/project.asp?ProjectID=358

UNICEF and UIS (2016). *Monitoring education participation: Framework for monitoring children and adolescents who are out of school or at risk of dropping out.* UNICEF Series on Education Participation and Dropout Prevention, Vol I. Geneva: UNICEF Regional Office for Central and Eastern Europe and the Commonwealth of Independent States.

University of Chicago Consortium on Chicago School Research. (2014). *Selecting effective indicators.* College Readiness Indicator Systems Resource Series. Seattle, WA: Bill & Melinda Gates Foundation.