# The Classification of Welsh Language Education:

# Primary Schools

## 1. Introduction

This report considers the categorisation / classification of Primary Schools regarding Welsh language education. Using a statistical technique called Cluster Analysis, it suggests that Primary Schools in Wales fit into three, four or five specific groups (or types) of language education.

A previous report entitled 'The Classification of Welsh Medium Education: Secondary Schools' discussed the classification of all Secondary Schools in Wales. That paper provides full technical details of the process of cluster analysis, its limitations and advantages. Please refer to that paper and its Appendices for full details of the procedures (e.g. aggregation and statistical).

## 2. The PLASC database

The data used for analyses derives from the 2017/2018 Pupil Level Annual School Census (PLASC). Ten PLASC Welsh language variables were used to classify 1261 Primary Schools. These variables are:

**Variable 1: TEACHER _TCTO**
% Teachers in a School (Full Time and Part Time) who are:
      **TC** Qualified teachers teaching Welsh as a first language *plus*
      **TO** Qualified teachers teaching other subjects through the medium of Welsh.

**Variable 2: TEACHER_ NW**
% Teachers in a School who are:
      **NW** Qualified and able to teach Welsh or through the medium of Welsh**, but not doing so.**

**Variable 3: PUPIL_FLUENT**
% Pupils in a School who are fluent in Welsh
*NB: The percentage calculation excludes 'information refused'.*

**Variable 4: PUPIL_NOTFLUENT**
% Pupils in a School who can speak Welsh but not fluently
*NB: The percentage calculation excludes 'information refused'.*

**Variable 5: WELSH_AT_HOME**
% Pupils in a School who can speak Welsh at home.

**Variable 6: STUDY_OF_WELSH_L1**
% Pupils in a School who are taught Welsh as a first language**.**

**Variable 7: CLASS_MAIN**
% Classes in a School where Welsh is the sole or **main** medium of instruction.

**Variable 8: CLASS_PART**
% Classes in a School where Welsh is used as a teaching medium for **part** of the curriculum (i.e. less than half).

**Variable 9: CLASS_L2**
% Classes in a School where Welsh is taught as a second language only.

**Variable 10:  Welsh Governors**
% Welsh Governors of a School who are Welsh speaking.


A PLASC variable that was *not* entered into the Cluster analyses was a School's self-classification as a Welsh medium, English medium or some kind of bilingual school using the categorization system in the October 2007 WAG paper: '*Defining schools according to Welsh medium provision*'.

 https://beta.gov.wales/sites/default/files/publications/2018-02/defining-schools-according-to-welsh-medium-provision.pdf

This variable is considered later in this paper for comparisons with the Clusters generated in this Project.


3. **Limitations of Dimensions**

Other variables might be included in future analyses if the relevant Welsh language data becomes available:

    (1)  Assessment data; Welsh language progress of pupils across Key Stages.
    (2)  *Derived variables* (e.g. Pupils not Speaking Welsh at Home but who are studying Welsh as a L1 / through the *medium* of Welsh).
    (3)  Languages of school communities. As it stands, apart from 'Welsh speaking Governor' there is no variable that locates if a School is placed in a more Welsh language community / catchment area or is placed in a more Anglicized area.
    (4)  Provision in Welsh for pupils with Additional Learning Needs (ALNs).
    (5)  Languages of learning / teaching / electronic resources / curriculum materials.
    (6)  School language ethos e.g. announcements, displays on walls in classrooms and corridors, playground language(s), languages of Teacher Assistants and Ancillary staff.

**4. Cluster Analysis**

The statistical procedure used for classification of Schools is Cluster Analysis. Appendix 2 of the Secondary School report contains a detailed and technical discussion of this technique (including its use in this Project), and its limitations.

Cluster Analysis takes each Primary school and finds out which group (cluster) they best fit using the 10 PLASC dimensions. A Dendrogram and other statistics are used to find the number of clusters that best fit the data (**see Appendix 1**). The K-Means procedure was used for the cluster analyses (see the Secondary School report for details) with one exception where a hierarchical analysis was preferable for interpretability (see later).

**5. How many Clusters?**

Various statistical tests and graphs of 'best fit' (see **Appendix 1**) plus interpretability suggest that either a three, four or a five cluster solution is appropriate for the Primary school data. Details follow.

A Dendrogram for this dataset of 10 variables and 1261 primary Schools suggests a 'best fit' of three, four or five clusters (see **Appendix 1**). A Scree graph (see **Appendix 1**) suggests five clusters (at the 'elbow bend'). Experience suggests that interpretation of cluster solutions is very important. That is, neither a dendrogram, various indices nor a scree graph tend to be unambiguous as to the optimal number of clusters. Thus, examining several solutions is sensible. For example, a six cluster solution was found to be inappropriate as one cluster contained only five schools.

A two cluster solution was too insensitive to 'Welsh' school differences, containing 171 'Welsh language' Primary schools and 1090 'English language' schools. The three Cluster Solution using a K-Means analysis, (see **Appendix 2**) has a dispersion of Welsh medium, Bilingual schools and English medium Schools that is too insensitive to language differences. For example, it places some Second Language Welsh schools in the same cluster as schools with classes having Welsh as the main medium. Despite the three cluster K Means solution going through 11 iterations of reclassification, the final solution lacks interpretability (**see Appendix 2**).

However, a three cluster solution was produced by the same hierarchical analysis that produces the Dendrogram. A three cluster hierarchical cluster analysis provided a superior solution (see below) and is easily interpretable.

In summary, various tests of 'best fit' and interpretability suggest a solution of either three, four or five clusters, and these are now examined across the 10 variables and 1261 schools.

6. **Results for the 1261 Primary Schools**

A table in **Appendix 3** shows to which cluster each School was allocated on three separate analyses: 3 clusters, 4 clusters and 5 clusters.

In **Appendix 3**, the final column (2007 Classification) provides a School's self-classification as a Welsh medium, English medium, English with significant Welsh, Transitional or Dual Stream School using the October 2007 WAG paper: '*Defining schools according to Welsh medium provision*'. The degree of matching between the 'old' 2007 classification and the findings of this paper is considered later.

To understand the difference between the clusters in each solution, the full data for each School is given in **Appendix 4**. The total number of schools in each cluster in the 3, 4 and 5 cluster solutions is given in the table below.

|  | No of Schools Welsh Medium | No of Schools English Medium | No of Schools Bilingual | No of L2 Schools with 'Progression' Potential | No of Schools Welsh Medium 2 |
|---|---|---|---|---|---|
| **3 Clusters** | 397 | 843 | 21 | n/a | n/a |
| **4 Clusters** | 385 | 752 | 22 | 102 | n/a |
| **5 Clusters** | 155 | 749 | 22 | 102 | 233 |

There follows an outline of the three, four and five cluster solutions in turn. The labels of the clusters are temporary and for discussion.

## Three, Four and Five Cluster Solutions

How do Primary schools in Wales fit into three, four or five groups (clusters)? Below is an overview of the choices using three tables: the three cluster solution, the four cluster solution and the five cluster solution. With the exception of the last line (Number of Schools in a Cluster), all the figures are percentages.

| 3 CLUSTER SOLUTION | Welsh Medium Schools | English Medium Schools | Bilingual Schools |
|---|---|---|---|
| %Teachers WELSH MEDIUM | **85** | 0 | 62 |
| %Teachers POTENTIAL WELSH | 1 | 3 | 1 |
| %PUPILS FLUENT in WELSH | **51** | 1 | 4 |
| %PUPILS PART FLUENT in WELSH | 32 | 23 | **37** |
| %PUPILS with WELSH at HOME | **43** | 1 | 14 |
| %PUPILS TAUGHT WELSH AS L1 | **95** | 0 | 6 |
| %CLASSES WELSH MAIN MEDIUM | **97** | 0 | 0 |
| %CLASSES WELSH PART MEDIUM | 0 | 0 | **100** |
| %CLASSES L2 WELSH | 3 | **100** | 0 |
| %WELSH GOVERNORS | **77** | 7 | 43 |
| **NUMBER OF SCHOOLS IN CLUSTER** | **397** | **843** | **21** |

In the above (and below) tables, the main differences between clusters are indicated in **bold**. In the three cluster solution, Welsh medium schools number 397 showing a higher percentage compared with the other two clusters on the following variables: teachers who are Welsh medium, pupils fluent in Welsh, pupils with Welsh at home and are taught Welsh as a first language. 97% of classes in this first cluster are Welsh medium and 77% of Governors are Welsh speaking. In contrast, 100% of English medium schools teach Welsh as a second language and 1% of pupils are fluent in Welsh.

The third cluster comprises 21 'Bilingual schools' who have 100% of classes where Welsh is a part medium: Ysgol Ein Harglwyddes, Ysgol Capelulo, Ysgol Deganwy, Ysgol Ffordd Dyffryn, Ysgol Glanwydden, Ysgol Talhaiarn, Ysgol Pant Y Rhedyn, Ysgol Craig Y Don, Ysgol Glan Conwy, Ysgol Babanod Llanfairfechan, Ysgol Porth Y Felin, Ysgol Bodafon, Ysgol San Sior, Ysgol Rhewl, Llanelwedd Church In Wales, Ysgol Gymunedol Cei Newydd, Ysgol Gymunedol Comins, Ysgol Gymunedol Plascru, Croesgoch CP, Holy Name RC School, Ysgol Gymunedol Llanedi. This Bilingual group of 21 school stays the same in the four and five cluster solutions except that Ysgol Gynradd Dolau is added in those solutions.

The four cluster solution (see the table below) has distinct Welsh medium and English medium clusters although there are slightly less schools in these clusters compared with the three cluster solution. A new cluster provisionally labelled 'Welsh Second Language with Progression Potential' numbers 102 schools. This new cluster is created by the movement of some schools from the Welsh medium and English medium clusters in the Three Cluster solution. This new cluster has the highest percentage (20%) of teachers who could teach through Welsh but are not doing so, and hence possibly gives potential progression to more Welsh medium teaching and learning. A list of schools in this 'Welsh Second Language with Progression Potential' cluster is given in **Appendix 5.**

| **4 CLUSTER SOLUTION** | **Bilingual Medium Schools** | **Welsh Medium Schools** | **Welsh L2 with Progression Potential** | **English Medium Schools** |
|---|---|---|---|---|
| %Teachers WELSH MEDIUM | 60 | **87** | 2 | 1 |
| %Teachers POTENTIAL WELSH | 1 | 1 | **20** | 1 |
| %PUPILS FLUENT in WELSH | 5 | **52** | 1 | 0 |
| %PUPILS PART FLUENT in WELSH | 35 | 32 | 27 | 23 |
| %PUPILS with WELSH at HOME | 15 | **44** | 3 | 1 |
| %PUPILS TAUGHT WELSH AS L1 | 8 | **97** | 3 | 0 |
| %CLASSES WELSH MAIN MEDIUM | 2 | **99** | 4 | 0 |
| %CLASSES WELSH PART MEDIUM | **98** | 0 | 0 | 0 |
| %CLASSES L2 WELSH | 0 | 1 | **95** | **99** |
| %WELSH GOVERNORS | 42 | **78** | 11 | 7 |
| **NUMBER OF SCHOOLS IN CLUSTER** | 22 | 385 | 102 | 752 |

The five cluster solution (see below) reproduces almost the same classification of schools as in the four cluster solution: English medium (749 schools), Bilingual (22 schools), and Welsh L2 with Progression Potential (102 schools). What is different from the four cluster solution is a split in the

Welsh medium schools. 'Welsh Medium 1' schools number 155 compared with 'Welsh Medium 2' with 233 schools. 'Welsh medium 1' schools have a much higher percentage of pupils who are fluent in Welsh, who speak Welsh at home and have relatively more Welsh speaking Governors. 'Welsh medium 2' schools have a comparatively high percentage of pupils who are partly fluent in Welsh. **Appendix 3** lists the schools in these two clusters.

| **5 CLUSTER SOLUTION** | Welsh Medium 1 | Welsh Medium 2 | Welsh L2 with Progression Potential | English Medium Schools | Bilingual Medium Schools |
|---|---|---|---|---|---|
| %Teachers WELSH MEDIUM | **88** | **85** | 2 | 0 | 60 |
| %Teachers POTENTIAL WELSH | 1 | 1 | **20** | 1 | 1 |
| %PUPILS FLUENT in WELSH | **73** | 38 | 1 | 0 | 5 |
| %PUPILS PART FLUENT in WELSH | 19 | **40** | **27** | 22 | 35 |
| %PUPILS with WELSH at HOME | **69** | 28 | 3 | 1 | 15 |
| %PUPILS TAUGHT WELSH AS L1 | **99** | **96** | 3 | 0 | 8 |
| %CLASSES WELSH MAIN MEDIUM | **100** | **97** | 4 | 0 | 2 |
| %CLASSES WELSH PART MEDIUM | 0 | 0 | 0 | 0 | **98** |
| %CLASSES L2 WELSH | 0 | 3 | **95** | 100 | 0 |
| %WELSH GOVERNORS | **92** | 69 | 11 | 6 | 42 |
| **NUMBER OF SCHOOLS IN CLUSTER** | **155** | **233** | **102** | **749** | **22** |

## Explaining the Results from the Cluster Solutions

The tables above suggest that there are two relatively stable clusters: English medium schools who number close to 750, and Bilingual schools who number 21 or 22. The English medium, Bilingual Medium and 'Welsh L2 with Progression Potential' are virtually the same in the four and five cluster solutions.

The Welsh medium schools number 397 in the three cluster solution, and 385 in the four cluster solution. These 385 schools (with three exceptions) are divided into two groups in the five cluster solution (155 'Welsh medium 1' Schools and 233 'Welsh medium 2' Schools). **Appendix 6** provides a numerical comparison of the four and five cluster solutions. The only real change from four to five clusters concerns the Welsh medium schools.

The difference in the two Welsh medium clusters found in the five cluster solution is shown in the table above using **bold** numerals and can be summarised as:

- Welsh medium Group 1 has 73% of pupils fluent in Welsh while Group 2 has 38%.
- Welsh medium Group 1 has 19% of pupils partly fluent in Welsh while Group 2 has 40%.
- Welsh medium Group 1 has 69% of pupils who come from Welsh speaking homes while Group 2 has 28%.
- Welsh medium Group 1 has 92% of Governors who are Welsh speaking while Group 2 has 69%.

- The two Groups have almost the same percentages for (i) percentage of teachers who teach Welsh or through Welsh, (ii) the percentage of pupils who are taught Welsh as a L1 and (iii) the percentage for whom Welsh is the main medium of teaching and learning.

**Conclusion**: There is a choice between a three, four or five cluster solution. It depends on a decision as to whether Welsh medium schools should be (i) split into two groups (i.e. a five cluster solution) (ii) combined (i.e. a in the three and four cluster solutions) or (iii) nesting the two Welsh groupings by creating a four group solution with a sub-division. The perceived value of a 'Welsh L2 with Progression Potential' cluster also determines which solution is preferred (e.g. the three of four cluster solutions).

## Comparing the 2007 WAG Classification system with the Four Cluster Solution

The current PLASC data includes a School's self-classification as Welsh medium, English medium, Dual Stream, English with some Welsh and Transitional using the October 2007 WAG paper: '*Defining schools according to Welsh medium provision*'.

How does this current PLASC classification compare with the clusters found in these analyses?

We are not comparing like with like. The WAG 2007 Secondary School classification system concerns 'text defined' dimensions of Welsh and English schools. In comparison, this analysis uses 10 PLASC dimensions. Therefore, we cannot expect 100% similarity.

One comparison is the 2007 WAG Classification used in PLASC 2017/18 and for this report compared with the Four Cluster Solution in the analyses presented in this paper. The result is **88.4% similarity**, with approximately one in every ten schools differently allocated across the two systems.

The results (**see below**) suggest that the classification of both Welsh medium and English medium schools is similar between the two systems with 'same classification' scores of **96.9%** for English medium schools, and **95.1%** for Welsh medium schools.

**Four Cluster Solution compared with the WAG 2007 Classification System**

| | | | Dual Stream | English Medium | English with some Welsh | Transitional | Welsh Medium | Total |
|---|---|---|---|---|---|---|---|---|
| **Four Cluster Solution** | Bilingual Medium | Count | 2 | 2 | 18 | 0 | 0 | 22 |
| | | % | 9.1% | 9.1% | **81.8%** | 0.0% | 0.0% | 100.0% |
| | Welsh Medium | Count | 15 | 0 | 0 | 4 | 366 | 385 |
| | | % | 3.9% | 0.0% | 0.0% | 1.0% | **95.1%** | 100.0% |
| | Welsh Teaching Potential | Count | 7 | 90 | 4 | 0 | 1 | 102 |
| | | % | 6.9% | **88.2%** | 3.9% | 0.0% | 1.0% | 100.0% |
| | English Medium | Count | 8 | 729 | 14 | 0 | 1 | 752 |
| | | % | 1.1% | **96.9%** | 1.9% | 0.0% | 0.1% | 100.0% |

| Totals | Count | 32 | 821 | 36 | 4 | 368 | 1261 |
|--------|-------|------|------|------|------|------|--------|
|        | %     | 2.5% | 65.1% | 2.9% | 0.3% | 29.2% | 100.0% |

The result suggests that:

(i) The classification of English medium schools is very similar between the two systems (96.9% similarity).

(ii) The Welsh Medium cluster has 95.1% similarity with the 2007 combined 'Bilingual' category.

(iii) The 'Bilingual' cluster has 81.8% similarity with the 2007 'English with some Welsh' category.

# END