Phonics screening check 2026

Technical specification, version 1.0

December 2025

Contents

Table of figures

Version history

Version	Change History	Author /Date
1.0	Dates rolled forward Links updated	Jon Leyshon 10 December 2025

1. Introduction

This specification covers the return of end of the phonics screening check to the Department for Education. These returns will be made using the department's COLLECT system, in XML format. It should be read in conjunction with the current Common Basic Dataset definitions and 2026 assessment components.

1.1. Audience and purpose

This specification describes the phonics screening check data collection 2026. It will enable schools, academies (including free schools), local authorities, and software suppliers working on their behalf, to prepare the necessary data and processes so that information on all phonics screening checks carried out in 2026 can be returned to the department.

The following legislation, as amended and currently in force, underpins the provision of this information:

- the Education Act 1996 s.537A(3)
- the Education (School Performance Information)(England) Regulations 2007

1.2. Scope

This data collection process covers all maintained schools and academies (including free schools) in England, submissions are processed through a local authority who in turn will provide data on all phonics screening checks carried out in 2026. The data consists of pupil-level information for all year 1 pupils and those pupils in year 2 who did not meet the expected level in the 2025 phonics screening check.

The following diagram outlines the overall scope of the collection:

Figure 1: high-level data collection process
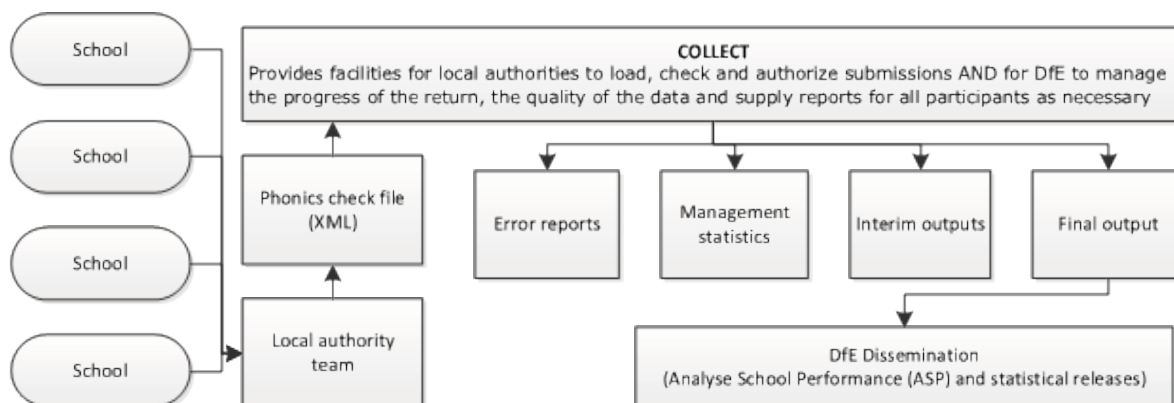
1.3. Changes for 2026

Changes for the 2026 phonics screening collection technical specification are:

- roll-over of dates
- updating of links

1.4. Coverage and timings

Maintained schools (including academies and free schools) will capture the data specified for the 2026 phonics screening check. The data will then be collected via department's COLLECT system between 8 June and 17 July 2026. Data received after 17 July 2026 will not be included in the provisional statistics. There will be a limited opportunity for LAs to submit late and / or corrected results but any late results will only feed into the National Pupil Database.

2. Structure and guidance

2.1. Outline data content

Each return from a local authority will consist of a header and one or more school modules (only one school per return as shown in Figure 2), each of which will contain a pupil module with a number of pupil records. Each pupil record will include identifiers plus the mark and outcome achieved in the phonics screening check.

Figure 2: outline structure of data
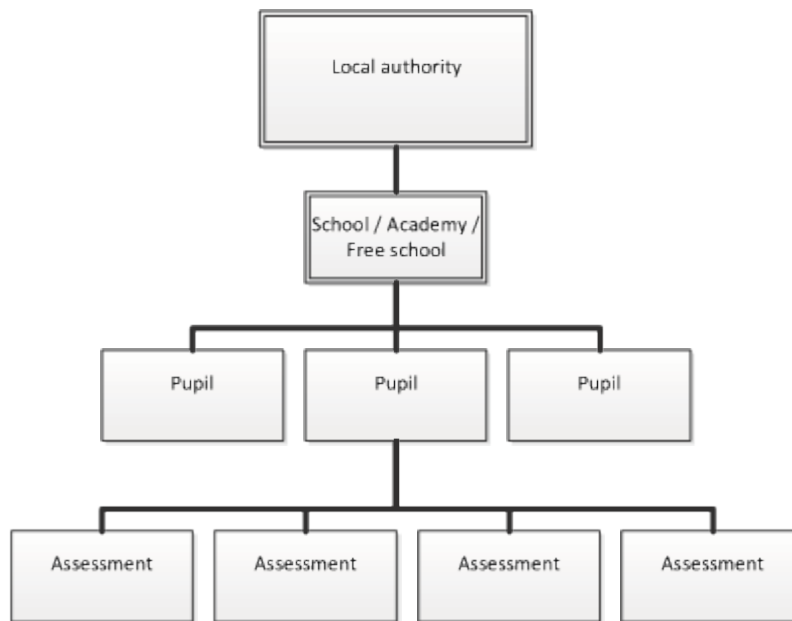
2.2. Process

The process will involve an annual screening check where teachers will assess students on their phonics decoding ability. Students will be asked to read out 40 items with a teacher. These items will be a mix of real words and pseudo-words. The teacher will then score each item (1 mark for a correct answer) and total the scores to make up the mark for the screening check. The threshold mark will be made available to schools after the end of the test window on Monday 22 June on GOV.UK. Schools are only required to submit the mark (or outcome where the pupil has not taken the test) to local authorities and local authorities will be required to submit this to the Department. The Department will apply the threshold after the data has been submitted to local authorities.

Local authorities will monitor 10% of the schools in their area. Monitoring visits will take place before the check (to check materials are being held securely), and during the time of the check to ensure the administration process is being followed and observe some screening checks taking place. If necessary, local authorities can also carry out monitoring visits immediately after the check week. Local authorities have been informed

of the general arrangements, and the funding being provided for this work. A guidance document will be sent out for monitors to local authorities in the spring term of 2026.

2.2.1. Entry of marks and outcomes into school MIS

Data will be gathered in the collection for all year 1 pupils and those in year 2 who did not meet the expected level in the 2025 phonics screening check. In addition to the basic pupil identifiers (as collected in other collections), there will be 2 data items for each pupil.

The two data items 'NC Year Group Actual' and 'Phonics Mark' are the required components of the 'Phonics Screening Check' data collection.

The 'Phonics Outcome' field is to be populated with the following values only where a pupil has not taken the test:

A – Absent

Q – Maladministration

D – Disapplied

L – Left

Outcomes Wa and Wt are not required. However, if these are included in a file returned to the department the file will be accepted and the Phonics Outcome field removed.

2.2.2. MIS data entry checks

The school MIS software should check that a valid entry has been made in the fields for the phonics screening check where appropriate. Any invalid or missing entries will be reported to the school when the data is entered and / or when the CTF is to be created.

2.2.3. Creation of CTF phonics file for transmission to local authority

Maintained schools and academies (including free schools) must submit the results of the phonics screening check to their local authority.

Academies (including free schools) must have a written agreement in place with their chosen local authority.

This will normally be via a CTF and when using this route, the name of the CTF created by schools will be:

xxxnnnn_PHO_xxxLLLL_yyy.XML where xxx represents the local authority number, nnnn is the school number and yyy is an incremented version number starting at 001.

2.2.4. Loading phonics results into the local authority's central management system

Where any pupil-identifying or contextual data are already present in the local authority central database there should be a prompt to ask if the existing data should be overwritten by the CTF or CSV import.

Where a CTF contains any pupil's previous assessment data (earlier year's foundation stage profile result), there should be a prompt to ask if the existing assessment data should be overwritten.

The local authority should also have facilities:

- to remove and to add pupils;
- to amend a pupil's identifying data or result levels that have been loaded, and for the new data to be re-validated in accordance with the rules in sections 4.1 to 4.3.
- to re-import a corrected CTF with an option for the operator to either replace all existing data or add as new data, and be validated in accordance with the rules in sections 4.1 to 4.3.

2.2.5. Creation of individual pupil results XML export file by the local authority for transmission to the DfE

On completion of the actions at 2.2.4, the software should allow the local authority to create separate XML files of individual school and pupil data, as specified in the example XML file at section 3.2, even if still containing errors, for loading into the department's data collection system (COLLECT). Where both phonics outcome and phonics mark are missing therefore no assessment result level is reported for a subject, all assessment components for that subject should be excluded from the XML file. The XML file will be validated in accordance with the rules in sections 4.1 to 4.3. Any data failing the validation checks will be reported using the error codes and messages specified. In addition to local authority and school number, the file will contain the following data items for each pupil:

- UPN
- pupil surname
- pupil forename
- DoB
- sex
- national curriculum year group
- phonics check result – mark as recorded
- phonics check result – outcome, if no marks as recorded. Outcome can be returned where a mark has been recorded but will be stripped out by COLLECT

Please note – An approved 'SoftwareCode' will not be issued. COLLECT will use the 'SoftwareCode' tag value specified by suppliers for this purpose.

The convention for naming the XML file is: xxxLLLL_PHO_xxxDfE_yyy.XML where xxx represents the local authority number and yyy is an incremented version number starting at 001.

The local authority should be provided with options to include in the export XML:

- all schools' files even if these contain errors or have previously been exported (- the software should inform the operator if a school is in error or has previously been exported)
- selected files (that is, a single school or multiple schools), even if these contain errors or have previously been exported (the software should inform the operator if a school is in error or has previously been exported)

Please note – Whichever option above is chosen, all individual pupil results within the school are to be exported. There must be one file per school.

2.2.6. Loading phonics results into the local authority's central management system.

In addition, the software should allow the local authority to load into its central database a CSV file created by schools from a phonics Excel spreadsheet (to be supplied to local authorities by the department for distribution if necessary) with the filename 'Phonics spreadsheet v1.0'. A copy of the spreadsheet can be provided when available, if required.

2.3. Additional local authority software facilities

The software should provide the following additional facilities:

• to allow the local authority to produce a list of schools to check that all expected schools' returns have been processed

• to allow the local authority to produce a report for each school or for all schools to show total number of pupils included in each school's CTF submission. It could also include number of boys and girls as an additional check

• to allow the local authority to transmit results in batches without the need to resubmit results already transmitted

2.4. Reports for schools and parents

Schools will be expected to report to parents or persons with parental responsibility, the phonics screening check outcomes for their children. Further information is given in the [2026 assessment and reporting arrangements (phonics screening check)](#)

3. Data return formats

A submission file from a local authority will be provided in an XML file structure as described in 3.2. Where an optional item is not included then the associated tag will also be omitted. Where a mandatory item is omitted then an error will be generated.

3.1. Special notes for XML returns

The standard XML entity references should be used for the following special characters:

Character	Entity reference
Ampersand (&)	&
Left angle bracket (<)	<
Right angle bracket (>)	>
Single quote / apostrophe (')	'
Double quotes (")	"

For special characters such as é use a character reference such as é This produces a generic XML file which can be viewed as such in a suitable browser. Details of all such characters are found within the XML standard documentation; see for example http://www.w3.org/TR/xml/, section 2.2.

3.2. Example file structure for XML returns

CBDS Ref	XML format layout	Notes
	<?xml version="1.0" encoding="UTF-8"?>	
	<PhonicsFile>	
	<Header>	
N00600	<Collection>Phonics Transfer File</Collection>	Fixed value of phonics transfer file
N00609	<DateTime>2026-05-23T15:30:47</DateTime>	
N00602	<Year>2026</Year>	Fixed value of '2026'
N00216	<LEA>555</LEA>	Valid values listed in Section 6
N00605	<SoftwareCode>KEYPAS</SoftwareCode>	Suppliers decide on code used by their system
	</Header>	
	<School>	
N00279	<Estab>1234</Estab>	Valid values listed in section 6
	<Pupils>	
	<Pupil>	A repeatable group of 1 or more
N00001	<UPN>A123456789012</UPN>	
N00003	<Surname>SURNAME 1</Surname>	

CBDS Ref	XML format layout	Notes
N00004	<Forename>FORENAME 1</Forename>	
N00006	<DOB>2019-02-02</DOB>	
N00783	<Sex>M</Sex>	
N00182	<NCyearActual>1</NCyearActual>	NCyearActual must be included and would be expected to be 1 or 2
	<Assessments>	
	<Assessment>	<Assessment> : A repeatable group of 1 or more depending on the combination of <Subject>, <Method>, <Component> and <ResultQualifier> required. See A_Comp 2026. Where a phonics mark is supplied there is no requirement for result to be returned and the xml structure should look as shown. Where no mark is to be returned the result can be returned as <Result>A</Result>. Results of Wa or Wt can be returned but will be stripped out by COLLECT
N00134	<Subject>PHO</Subject>	
N00062	<Method>TT</Method>	
N00063	<Component>CHK</Component>	
N00200	<ResultQualifier>NY</ResultQualifier>	
N00064	<Result/>	
	</Assessment>	

CBDS Ref	XML format layout	Notes
	<Assessment>	<Assessment> : A repeatable group of 1 or more depending on the combination of <Subject>, <Method>, <Component> and <ResultQualifier> required. See A_Comp 2026.
N00134	<Subject>PHO</Subject>	
N00062	<Method>TT</Method>	
N00063	<Component>CHK</Component>	
N00200	<ResultQualifier>NM</ResultQualifier>	
N00064	<Result>39</Result>	
	</Assessment>	
	</Assessments>	
	</Pupil>	
	</Pupils>	
	</School>	
	</PhonicsFile>	

4. Data validation rules

This section specifies validation rules that are applied to the data by Department for Education, using COLLECT. They must also be applied, as far as possible, by local authorities prior to loading the data onto the COLLECT system and any issues resolved. Those data items included in the specification but not appearing in the validation rules - that is, those not shown in the validation rules as "must be present" - must be supplied where available (for example, child's unique pupil number). The rules also show which validations are errors and which ones are queries. An error is a failed validation check that must be corrected. A query is one which must be investigated, and potentially corrected, as it identifies data that would usually be invalid.

Syntax:

- Validation checks are expressed as conditions that must be true (errors) or that should usually be true (queries). The Error Message shows what is displayed if the condition is not met.
- data items are indicated using the XML tag with the common basic data set number in parentheses, for example <UPN>
- data groups are indicated using the XML tag only, for example <Assessments>
- single dates or date ranges are sometimes defined in relation to <ReferenceDate>, which is from the XML Header.

4.1. Header validation

Seq Number	Error/ Query	Validation Check	Error Message
001	Error	<Collection> must be present with a value of "Phonics Transfer File"	Document Name missing or invalid. The file must be recreated
101	Error	<SoftwareCode> must be present	Software Code is missing
110	Error	<Year> must be present and = 2026	Year is missing or invalid

4.2. Establishment level validation

Seq Number	Error/ Query	Validation Check	Error Message
116Q	Query	COLLECT ONLY CHECK a query will be raised where the number of boys where <NCyearActual> = 1 is fewer than the expected number of boys by more than 10	The number of boys in NC year 1 loaded is significantly different to the number of boys expected. (expected=xx and loaded =xx)
117Q	Query	COLLECT ONLY CHECK a query will be raised where the number of girls where <NCyearActual> = 1 is fewer than the expected number of girls by more than 10	The number of girls in NC year 1 loaded is significantly different to the number of girls expected (expected=xx and loaded =xx)
118Q	Query	The sum of <NCyearActual> pupils in year 2 is greater than 0 and the sum of <NCyearActual> pupils in year 1 = 0	Please check: only NC year 2 pupils reported

Seq Number	Error/ Query	Validation Check	Error Message
119Q	Query	COLLECT ONLY CHECK a query will be raised where the number of boys where \<NCyearActual\> = 2 is fewer than the expected number of boys by more than 10	The number of boys in NC year 2 loaded is significantly different to the number of boys expected (expected=xx and loaded =xx)
120Q	Query	"COLLECT ONLY CHECK a query will be raised where the number of girls where \<NCyearActual\> = 2 is fewer than the expected number of girls by more than 10"	The number of girls in NC year 2 loaded is significantly different to the number of girls expected (expected=xx and loaded =xx)
121Q	Query	The sum of \<NCyearActual\> pupils in NC year 1 should be an integer value greater than zero	Please check: No NC year 1 pupils have been reported
122Q	Query	The sum of \<NCyearActual\> pupils in NC year 2 should be an integer value greater than zero	Please check: No NC year 2 pupils have been reported
301	Error	\<LEA\> must be provided and must be a valid value	Local Authority number missing or invalid
302	Error	\<Estab\> must be present with a valid value	Establishment No is missing
305	Query	The sum of \<NCyearActual\> pupils in year 2 is greater than the sum of \<NCyearActual\> pupils in year 1	Please check: more NC year 2 pupils reported than NC year 1 pupils

4.3. Pupil level validation: Pupil Identifiers

Seq Number	Error/ Query	Validation Check	Error Message
1600	Error	<DOB> must be present and in a valid format	Pupil's date of birth is missing or in an invalid format. Format should be CCYY – MM - DD
1601Q	Query	Where <NCyearActual> is equal to '1', then <DOB> should be between 1/9/2018 and 31/8/2021 OR Where <NCyearActual> is equal to '2', then <DOB> should be between 1/9/2017 and 31/8/2020	Pupil's Date of Birth is outside expected date range
1610	Error	<Sex> must be present and a valid value	Pupil sex missing or invalid
1500	Error	<UPN> (Unique Pupil Number) must be present	UPN is missing

Seq Number	Error/ Query	Validation Check	Error Message
1510	Error	If UPN is present then <UPN> must contain the correct check letter To calculate the check letter: 1. Multiply the individual digits by their weights as follows: digit 2 by weight 2; digit 3 by weight 3; digit 4 by weight 4; digit 5 by weight 5; digit 6 by weight 6; digit 7 by weight 7; digit 8 by weight 8; digit 9 by weight 9; digit 10 by weight 10; digit 11 by weight 11; digit 12 by weight 12; digit 13 by weight 13. 2. Sum the individual results, divide the total by 23, and take the remainder. 3. Calculate the check letter from the result as follows: 0 = A; 1 = B; 2 = C; 3 = D; 4 = E; 5 = F; 6 = G; 7 = H; 8 = J; 9 = K; 10 = L; 11 = M; 12 = N; 13 = P; 14 = Q; 15 = R; 16 = T; 17 = U; 18 = V; 19 = W; 20 = X; 21 = Y; 22 = Z. For calculating the check letter (see above) any alphabetical character at digit 13 is accorded the same numerical value as listed in paragraph 3 above, A=0, B=1, C=2 etc.	UPN invalid (wrong check letter at character 1)
1520	Error	Each pupil <UPN> must be unique across all pupils in the Estab	More than one pupil record with the same UPN
1530	Error	If <UPN> is present Characters 2-4 of <UPN> must be a valid post April 1999 LA code or a recognised "pseudo LA" code (001-005, 201-213, 301-320, 330-336, 340-344, 350-359, 370-373, 380-384, 390-394, 420, 660-681, 701-708, 800-803, 805-808, 810-813, 815, 816, 820-823, 825, 826, 830, 831, 835-837, 838-839, 840, 841, 845, 846, 850-852, 855-857, 860, 861, 865-896, 908, 909, 916, 919, 921, 925, 926, 928, 929, 931, 933, 935-38, 940-943)	UPN invalid (characters 2-4 not a recognised LA code)

Seq Number	Error/ Query	Validation Check	Error Message
1540	Error	If <UPN> is present Characters 5-12 of <UPN> must be numeric	UPN invalid (characters 5-12 not all numeric)
1550	Error	If <UPN> is present Characters 13 of <UPN> must be numeric or A-Z omitting I, O and S	UPN invalid (character 13 not a recognised value)
1580	Error	<Surname> must be present	Pupil with surname missing
1590	Error	<Forename> must be present	Pupil with forename missing
113	Error	Each <Assessment> should only appear once for each <Pupil> for the same combination of <Subject><Component><Method><ResultQualifier>	Assessments are duplicated
124	Error	For each pupil there must be one and only one instance where <Subject>=PHO and <Component>=CHK and <Method> =TT and <ResultQualifier>=NY	There are multiple assessments where the Subject, Component, Method and Result Qualifier combinations are the same.
125	Error	If there is no assessment where <ResultQualifier>=NM and <Result>= 0 - 40 then there must be an assessment where <Subject>=PHO, <Component>=CHK <ResultQualifier>=NY and <Result> is one of A, Q, D, or L for the same pupil	The outcome result is invalid and should be one of A, Q, D, or L
126	Error	For each pupil there must be no more than one instance where <Subject>=PHO and <Component>=CHK and <Method> =TT and <ResultQualifier>=NM	There are multiple assessments where the Subject, Component, Method and Result Qualifier combinations are the same.
127	Error	If <Subject>=PHO, <Component>=CHK and <ResultQualifier>=NM then <Result> must be an integer in the range 0 – 40 inclusive	The outcome result is invalid and must be in the range 0 to 40

Seq Number	Error/ Query	Validation Check	Error Message
129	Error	If <ResultQualifier>=NM and its result is present AND there is a record for the same pupil where <ResultQualifier>=NY then <Result> must not be A,Q,D or L	Phonics outcome and mark invalid
130	Error	Each <Assessment> must contain <Subject>=PHO	There is an unexpected entry in the Subject container
131	Error	Each <Assessment> must contain <Component>=CHK	There is an unexpected entry in the Component container
132	Error	Each <Assessment> must contain <Method>=TT	There is an unexpected entry in the Method container
133	Error	For each <Pupil> record at least one <Assessment> node must be included.	No assessment data is included for this pupil
134Q	Query	Number of <Assessment> nodes where <Result> equals A, divided by total number of <Pupil> nodes, should not be greater than 0.1.	Please check: proportion of absent pupils is above 10%
136Q	Query	There should be no <Assessment> nodes where <Result> equals Q.	Please check: One or more records marked as Q (maladministration).
139	Error	In each <Pupil> record <NCyearActual > must be present	National Curriculum year group not provided
141	Error	<NCyearActual > must be equal to '1' or '2'	National Curriculum year group must be either '1' or '2'.

Seq Number	Error/ Query	Validation Check	Error Message
142	Error	If <ResultQualifier>=NY then <Result> must be one of A, Q, D, L, Wa, Wt or BLANK	The outcome result is invalid and should be one of A, Q, D, L, Wa, Wt or 'blank'

5. Resubmission guidance

When resubmissions are required, the Serial No in the header **must** be incremented by the provider's software. We have also assumed that the whole file will be resubmitted. Note that a resubmission will **delete and replace all the data** contained in the previous submission.

6. Valid values

Valid item values must be in the list or range specified in CBDS, the [Common Basic Data Set](#).

6.1. Local authority codes

Local authority codes for this collection only covers ENGLAND local authorities.

Coverage also excludes local authority codes:

- 001, 003, 167
- codes starting with 6 and 7 except 702 (MoD schools, which is valid)
- XXX, MMM and NA

6.2. Establishment types

These are the following;

- maintained schools with a year 1 or year 2 cohort
- academies (including free schools)
- service children's education schools