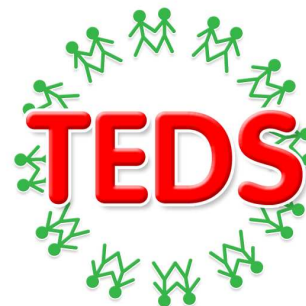


**Welcome to the TEDS SRCD Monograph dataset.**

**Before you use the data please take note of the following:**



## **1. Twin Identifiers**

Within the dataset, each row of data relates to one individual twin, uniquely identified by the "id\_twin" variable. (See example below.) An individual twin can also be specified by a combination of the "ID\_fam" variable (the family identifier) and the "twin" variable (the twin's birth order within the pair: 1=elder, 2=younger). Any pair of twins will share the same value for "ID\_fam", but will have different values for "id\_twin" and "twin".

In order to protect the confidentiality of the TEDS families, the IDs have been anonymised so that a given value of "ID\_fam" or "id\_twin" cannot be linked to an identifiable family or twin. However, the uniqueness of the identifiers has been preserved, as described above.

## **2. Double Entry of Twin Data**

Each individual twin has a row of data in the dataset, but that row contains the co-twin's data in addition to the individual twin's data. As a result, every item of twin-specific data is effectively duplicated within the dataset, and we say that the data are "double entered". Double entry is used in twin datasets in order to correlate one twin's data with the co-twin's data.

For example, variables "sex1" and "sex2" denote the sexes of the two twins in a pair. The value of "sex1" denotes the sex of the twin in the current row of data (identified by "id\_twin"). The value of "sex2" denotes the sex of the co-twin. Note the values of these two variables for the pair of twins with "id\_twin" values 11091 and 11092 (see screen shot below). 11091 is the elder of the two twins ("twin"=1), and 11092 is the younger of the two twins ("twin"=2). For the twin with "id\_twin"=11091, "sex1"=1, meaning that

this twin is male, while "sex2"=0, meaning that the co-twin is female. These values of sex1/sex2 are reversed in the row of data with "id\_twin"=11092.

	id_twin	ID_fam	twin	random	sex1	sex2	zygos	sexzyg
5	11091.00	1109.00	1	1	1	0	2	5.00
6	11092.00	1109.00	2	0	0	1	2	5.00

Note that in all cases of twin-specific data, the variable names end in "1" or "2". Any variable whose name ends in "1" relates to the twin identified by "id\_twin" in the current row of data. The corresponding variable whose name ends in "2" relates to the co-twin.

### 3. Selecting a Single Entered Dataset

Some analyses of twin data may require a sub-sample in which only one twin is selected from each pair. We refer to a selected sub-sample of this sort as a "single-entered" dataset.

If you select or filter cases in the dataset using the criterion "twin"=1, you will select one twin from each pair, but in every case the selected twin will be the elder twin of the pair.

For analyses of this sort, it is generally preferable to select a random distribution of elder and younger twins. The "random" variable is included in the dataset for this purpose. If you select or filter cases in the dataset using the criterion "random"=1, you will select one twin from each pair, but with a random selection of elder and younger twins. For any given twin pair (with the same value of "ID\_fam"), the values 0 and 1 have been randomly assigned to the "random" variable for the two individual twins.

### 4. Variable Names

TEDS uses an alphabetical system to identify the variables, for example, a prefix of "a" referencing data at age 1, "b" at age 2, "c" at age 3 etc.

Therefore age 7 raw data have the prefix “g”, age 9 raw data the prefix “i” and age 10 raw data the prefix “j”.

For example, the variables for National Curriculum English speaking and listening (‘esl’) for twin 1 at 7, 9 and 10 are:

*gesl1 iesl1 jesl1*

Age and sex corrected variables follow a similar pattern, all starting with the prefix “r” (for regressed) and followed by the letter to identify the age.

Following the above example, age and sex regressed National Curriculum English speaking and listening variables are:

*rgesl1 riesl1 rjesl1*

All variables are also comprehensively labeled in the dataset. (In SPSS, these labels can be found in *Variable View*.)

## **5. Exclusions**

We use exclusion variables to remove individuals from the dataset before analysis. Those individuals who meet the exclusion criteria at every age have already been removed from the dataset.

Please see chapter 2 for a description of the exclusion criteria.

In addition some individuals meet exclusion criteria at only one age. We have provided the specific age exclusion variables (coded as 1: Yes - exclude from dataset, and 0: No - keep in dataset).

These specific age exclusions should be used when analysis is done at just one age.

The specific age exclusion variables at 7, 9 and 10 are:

*gexclude*      *iexclude*      *jexclude*

## 6. Age Variables

Because teacher assessments and test data are not necessarily collected at the same time, we have recorded the age of the twins for both types of assessment.

This means that at every assessment year we have recorded the age of the twins when the teacher questionnaires were returned, and also the age when the test data were received.

For example the age variables for the 7 year data are:

*gtqage1* (g=7, t=teacher, q=questionnaire)

*gciage1* (g=7, ci=child interview)

As noted above, the dataset includes both 'raw' data and data corrected for age and sex (identified with the prefix 'r'). The age used for correction differed for teacher questionnaires and test data.

## 7. Outliers

Outliers have not been excluded. In the Monograph, individuals who scored +/- 3 SD were excluded from analyses of individual differences in the total sample. (but not the extremes analyses) so that individuals as rare as one in a thousand did not unduly influence the results. (See Chapter 2 for details.)

## 8. DF Extremes Analysis

As indicated in Chapter 2, DF extremes analysis uses variables corrected for age alone, not age and sex. Variables are only age corrected in order not to affect the representativeness of groups at low ability cut-offs. Rather than adding a third set of variables for age corrected variables to the raw data and the data corrected for age and sex, it will be necessary to create age corrected variables for DF analysis.

## 9. Zygosity Variables

Three zygosity variables are included:

- “zygos” a 2-value zygosity variable denoting either MZ or DZ pairs. (1=MZ; 2=DZ).
- “sexzyg” a 5-value zygosity variable denoting zygosity by sex. (1=MZ male; 2=DZ male; 3=MZ female; 4=DZ female; 5= DZ opposite-sex).
- “x3zygos” a 3-value zygosity variable denoting MZ, same-sex DZ, and opposite-sex DZ pairs. (1=MZ; 2=DZ same-sex; 3=DZ opposite-sex).

	Name	Type	Width	Decimals	Label
7	zygos	Numeric	1	0	Zygosity (best available estimate), 1MZ 2DZ
8	sexzyg	Numeric	8	2	Zygosity and sex, 1MZM 2DZM 3MZF 4DZF 5DZO 7unknown
9	x3zygos	Numeric	8	2	Three-value zygosity and sex, 1=MZ 2=DZSS 3=DZOS