Data Description

The dataset is aggregated from 8 different school-level surveys that conducted in rural areas of seven provinces of in 2014 and 2015, which covered 18888 students of grade 4 and 5 across 67 counties in rural China. The dataset includes information on student characteristics (gender, age, minority, whether the student boards at school) and family background (household size, parent education, and parent migration status) as well as academic performance (math test items for the test were selected from the standardized mathematics curricula for primary school students in China, and the validity of the test content was checked by multiple experts).

Codebook

---

id Identifier

---

type: string (str20)

unique values: 18888 missing "": 0/18888

examples: "1056127NCC_bs"
"2010233124SIL3_f1"
"2060713102SIL3_f1"
"3086127NCC_bs"

---

project Project - name

---

type: string (str5)

unique values: 7 missing "": 0/18888

tabulation: Freq. Value
2745 "NCC"
6197 "READ"
2931 "SIL3"
3659 "SL"

```
                 594  "TFF1"
                 757  "TFF2"
                2005  "Tp4p2"
```

---

project_n Project - number

```
        type:  numeric (float)

       range:  [1,8]                units:  1
unique values:  7                   missing .:  0/18888

  tabulation:  Freq.  Value
               6197  1
               2931  2
               3659  3
                594  4
                757  5
               2005  7
               2745  8
```

---

datawave Data wave - name

```
        type:  string (str8)

unique values:  8                   missing "":  0/18888

  tabulation:  Freq.  Value
               2745  "NCC_bs"
                656  "READ1_f1"
               5541  "READ2_bs"
               2931  "SIL3_f1"
               3659  "SL_bs"
                594  "TFF1_f1"
                757  "TFF2_bs"
               2005  "Tp4p2_f1"
```

---

datawave_n Data wave - number

---

type: numeric (long)
label: project_n

range: [1,10] units: 1
unique values: 8 missing .: 0/18888

tabulation: Freq. Numeric Label
2745 1 NCC_bs
656 2 READ1_f1
5541 3 READ2_bs
2931 4 SIL3_f1
3659 5 SL_bs
594 7 TFF1_f1
757 8 TFF2_bs
2005 10 Tp4p2_f1

---

date Date of survey round

---

type: string (str7)

unique values: 7 missing "": 0/18888

tabulation: Freq. Value
594 "01/2015"
2931 "05/2014"
3659 "05/2015"
757 "09/2015"
2745 "10/2015"
6197 "4/2015"
2005 "5/2015"

---

citycode citycode

---

type: numeric (float)

range: [1,19] units: 1
unique values: 18 missing .: 0/18888

mean: 11.2677
std. dev: 6.48813

percentiles: 10% 25% 50% 75% 90%
1 6 10 19 19

---

cityname cityname

---

type: string (str13)

unique values: 18 missing "": 0/18888

examples: "Ganzhou"
"Ganzhou"
"Longnan"
"Weinan"

---

countycode countycode - original project

---

type: numeric (float)

range: [101,207] units: 1
unique values: 17 missing .: 15363/18888

mean: 158.457
std. dev: 48.7247

percentiles: 10% 25% 50% 75% 90%
103 106 201 203 206

---

countyname County name

---

type: string (str13)

unique values: 67 missing "": 0/18888

examples: "Huichang"
"Nanzhen"
"Shiquan"
"Yudu"

---

provincecode provincecode

---

type: numeric (float)

range: [1,7] units: 1
unique values: 7 missing .: 0/18888

tabulation: Freq. Value
3779 1
4339 2
769 3
1059 4
656 5
5541 6
2745 7

---

provincename provincename

---

type: string (str7)

unique values: 7 missing "": 0/18888

tabulation: Freq. Value
4339 "Gansu"
656 "Guizhou"
5541 "Jiangxi"
1059 "Ningxia"

```
                 769  "Qinghai"
                3779  "Shaanxi"
                2745  "Sichuan"

-------------------------------------------------------------------------------------------
st_age                                                              Age (months)
-------------------------------------------------------------------------------------------

            type:  numeric (float)

           range:  [7,249]                  units:  1
      unique values:  144                   missing .:  0/18888

            mean:   131.503
          std. dev:   26.3602

       percentiles:        10%       25%       50%       75%       90%
                           118       125       133       144       153

-------------------------------------------------------------------------------------------
st_grade                                                            School grade
-------------------------------------------------------------------------------------------

            type:  numeric (float)

           range:  [4,6]                    units:  1
      unique values:  3                     missing .:  0/18888

        tabulation:  Freq.  Value
                     6455  4
                    10775  5
                     1658  6

-------------------------------------------------------------------------------------------
st_ethnic                                                 Ethnic minority (1=yes;0=no)
-------------------------------------------------------------------------------------------

            type:  numeric (float)

           range:  [0,1]                    units:  1
      unique values:  2                     missing .:  9787/18888
```

tabulation: Freq. Value
4971 0
4130 1
9787 .

---

st_male Gender (1=male;0=female)

---

type: numeric (double)

range: [0,1] units: 1
unique values: 2 missing .: 0/18888

tabulation: Freq. Value
9121 0
9767 1

---

st_board Student is boarding (1=yes;0=no)

---

type: numeric (double)

range: [0,1] units: 1
unique values: 2 missing .: 0/18888

tabulation: Freq. Value
13463 0
5425 1

---

hh_migr_dad Father is migrant (1=yes;0=no)

---

type: numeric (float)

range: [0,2] units: 1
unique values: 3 missing .: 3214/18888

tabulation: Freq. Value
8631 0
7041 1
2 2
3214 .

---

hh_migr_mom Mother is migrant (1=yes;0=no)

---

type: numeric (float)

range: [0,2] units: 1
unique values: 3 missing .: 3226/18888

tabulation: Freq. Value
10980 0
4673 1
9 2
3226 .

---

test_math_type Test type - Math (REAP test)

---

type: string (str2)

unique values: 2 missing "": 0/18888

tabulation: Freq. Value
8299 "M2"
10589 "M3"

---

std_test_math_score Standardized test math score

---

type: numeric (float)

range: [-3.3982656,2.3033779] units: 1.000e-09
unique values: 31 missing .: 0/18888

```
            mean:   9.2e-09
         std. dev:       1

      percentiles:       10%       25%       50%       75%       90%
                  -1.30766  -.737499    .02272    .78294    1.3531
```

---

std_math_score_m2 Standardized test math score for grade 4

```
            type:  numeric (float)

           range:  [-3.4832425,2.2710447]       units:  1.000e-09
      unique values:  31                   missing .:  10589/18888

            mean:   3.3e-09
         std. dev:       1

      percentiles:       10%       25%       50%       75%       90%
                  -1.37334  -.606099   .161139   .736568     1.312
```

---

std_math_score_m3 Standardized test math score for grade 4

```
            type:  numeric (float)

           range:  [-3.3392801,2.3328733]       units:  1.000e-08
      unique values:  31                   missing .:  8299/18888

            mean:   1.6e-09
         std. dev:       1

      percentiles:       10%       25%       50%       75%       90%
                  -1.25949  -.692275   .064012   .631227   1.19844
```

---

test_grade_begin Test score - Test on the beginning of the grade (1=begin 0=end)

---

type: numeric (float)

range: [0,1] units: 1
unique values: 2 missing .: 0/18888

tabulation: Freq. Value
15386 0
3502 1

---

cyrinc_new County per capita rural net income, thousands of yuan

---

type: numeric (float)

range: [3.234,13.086] units: 1.000e-07
unique values: 67 missing .: 0/18888

mean: 6.87275
std. dev: 1.90699

percentiles: 10% 25% 50% 75% 90%
4.981 5.96 6.792 7.156 9.22

---

m3county County average grade 5 standard students score

---

type: numeric (float)

range: [-1.0845549,1.0705411] units: 1.000e-09
unique values: 63 missing .: 8299/18888

mean: 3.8e-09
std. dev: .357509

percentiles: 10% 25% 50% 75% 90%
-.411557 -.316833 -.046855 .265282 .396581

---

m2county County average grade 4 standard students score

---

type: numeric (float)

range: [-1.0568514,1.2160922] units: 1.000e-10
unique values: 62 missing .: 10589/18888

mean: 3.8e-09
std. dev: .354004

percentiles: 10% 25% 50% 75% 90%
-.336944 -.3024 -.020741 .167218 .588715

---

scorecountyaver County average score

---

type: numeric (float)

range: [-1.0845549,1.2160922] units: 1.000e-10
unique values: 125 missing .: 0/18888

mean: 3.8e-09
std. dev: .355964

percentiles: 10% 25% 50% 75% 90%
-.405863 -.314122 -.033641 .189227 .447197