# Selection of data and metadata: balance and representativity

https://spraakbanken.github.io/swell-release-v1/Metadata-SweLL

SweLL release August 26-27, 2021


obrigado
Dank U
Merci
mahalo
Köszi
спасибо
Grazie
Thank you
maurururu
Takk
Gracias
Dziękuję
Děkuju
danke
Kiitos

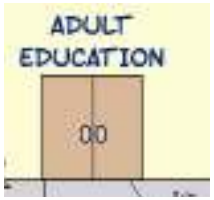
## Essays




What is task?






ADULT
EDUCATION







# Collection of data and metadata

- Collected 2017-2020 from north to south
- 451 adult learners of Swedish as a second language
- Authentic classroom/test situations



- 694 texts collected, transcribed and pseudonymized
- 502 texts also normalized and correction annotated



# Essay data (502 texts, T)



361 written by hand and transcribed



141 written on computer

**Purpose**

| | |
|---|---|
| Placement tests | 63 T |
| Final test *(incl. Tisus)* | 186 T |
| Formative *in the classroom* | 236 T |
| N/A | 17 T |

**Essay metadata**

Stockholm University

# Tasks and text types

- Text type
- Purpose
- Subject / title
- Task instructions & materials

- Time allowed
- Allowed materials (help)
- Grading scale
- Format (handwritten / digital)
- Date (for longitudinal students)

- Course books (optional)

| **Text type** | **Subject/topic 44 tasks e.g.** | **# essays** |
|---|---|---|
| Argumentative | *Behöver man en egen bil?* | 214 |
| Descriptive | *Berätta hur du bor!* | 104 |
| Reasoning, Expository | *Brott – orsaker och konsekvenser* | 77 |
| Narrative | *Min första kärlek* | 56 |
| Instructional | *Ge tips och råd - anställningsintervju* | 25 |
| Summarizing | *Referat av texten xxx* | 25 |
| **Total Gold** | | **502** |

| # students | # essays per student | Total # essays |
|---|---|---|
| 194 | 1 | 194 |
| 257 | 2-5 | 308 |



# Learner questionnaire
# 451 adult learners

**Personlig information** *(Var vänlig, bokstavera!)*

SweLL-ID _____

Kön: ☐ Kvinna ☐ Man ☐ Vill inte säga / annat

Född (år): ☐ 2000-2004 ☐ 1995-1999 ☐ 1990-1994 ☐ 1985-1989 ☐ 1980-1984
☐ 1975-1979 ☐ 1970-1974 ☐ 1965-1969 ☐ 1960-1964 ☐ 1955-1959
☐ 1950-1954 ☐ 1945-1949 ☐ 1940-1944 ☐ tidigare

Första språk (*ett eller flera*) _____

Total tid i Sverige _____ år _____ månader

**Utbildningsnivå**

| | *utanför Sverige* | *i Sverige* |
|---|---|---|
| ☐ Grundskola | antal år _____ | antal år _____ |
| ☐ Gymnasiet | antal år _____ | antal år _____ |
| ☐ Högskola / universitet | antal år _____ | antal år _____ |
| ☐ Annat _____ | antal år _____ | antal år _____ |

☐ Högsta examen _____

**Språk**

☐ Undervisning i svenska / svenska som andraspråk / SFI antal år ____ månader ____

☐ Undervisning i modersmål (*vilket språk?*) _____ antal år ____ månader ____

☐ Andra språk du kan: _____

☐ Vilket språk talar du bäst? _____

☐ Vilket språk skriver du bäst? _____

☐ Vilket/vilka språk använder du:

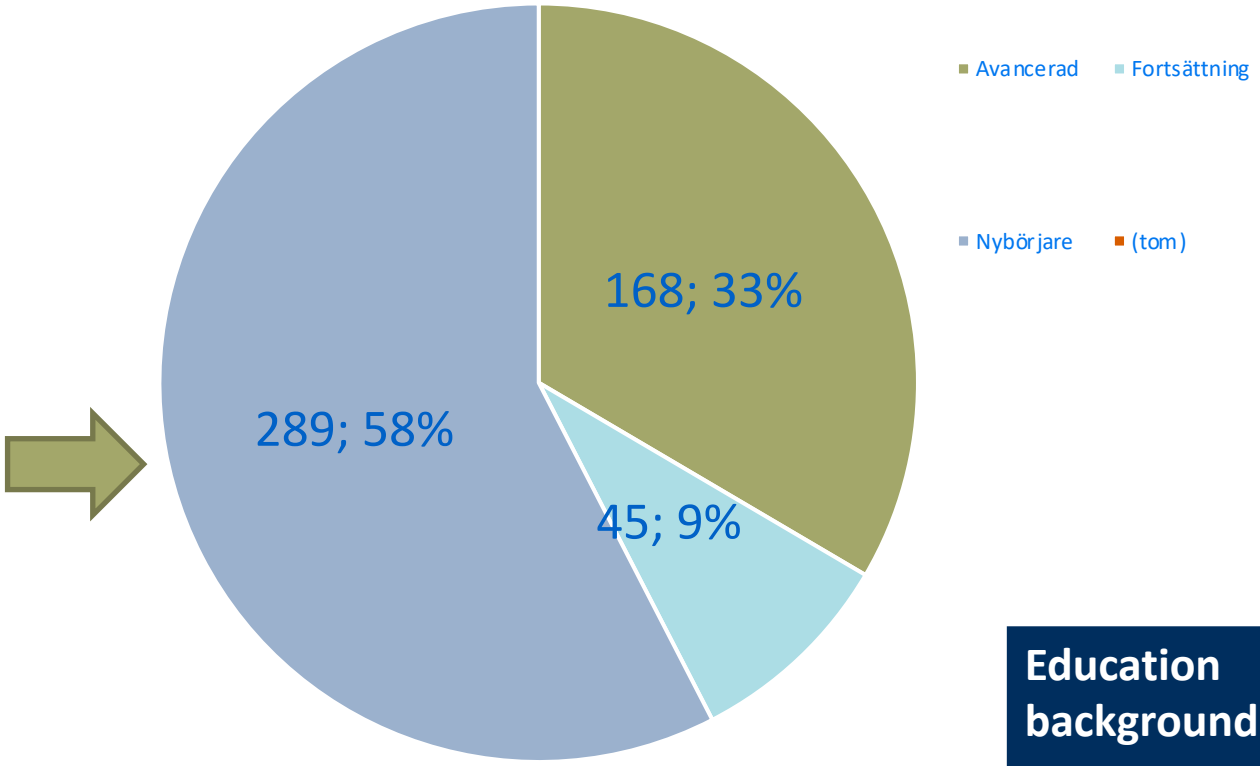
med din familj? _____

med vänner ?

# Courses of Swedish as a second language

| ADULT EDUCATION | SweLL # | Gymnasiet (16-19 years) | SweLL # | UNIVERSITIES | SweLL # | SweLL 502 (694) |
|---|---|---|---|---|---|---|
| **NYBÖRJARE/BEGINNERS** | | | | | | |
| **Svenska för invandrare** *167 511 participants* | | Språk-introduktion | | | | **289** (331) |
| Sfi A | I-prov 24 | | | | | |
| Sfi B | 16 | | | | | |
| Sfi C | 26 | | | | | |
| Sfi D | 223 | | | | | |
| **FORTSÄTTNING/ INTERMEDIATE** | | | | | | |
| **Grundläggande vux** *48 585* | | Språk-introduktion | | | | **45** (45) |
| Nationell delkurs 1 | ↓ | | | | | |
| Nationell delkurs 2 | 15 | | | | | |
| Nationell delkurs 3 | ↓ | | | | | |
| Nationell delkurs 4 | 30 | | | | | |
| **AVANCERAD/ADVANCED** | | | | | | |
| **Gymnasial vux** *56 787* | | **Gymnasiet** | | **Högskolan** | | **168** (318) |
| Sva 1 | | Sva 1 | | Förberedande kurs | I-prov 39 | |
| Sva 2 | | Sva 2 | 19 | Förberedande kurs | 39 | |
| Sva 3 | | Sva 3 | 29 | Behörighetsgivande k<br>Tisus | 27<br>15 | |

## Proficiency levels




## Learner metadata

| Education background | Learner texts (502) |
|---|---|
| 1. 0-6 years | 34 |
| 2. 7-9 years | 56 |
| 3. 10-13 years | 155 |
| 4. 14- years | 257 |

## Age

| Age | Learner texts (502) * |
|---|---|
| 16-20 years old | 37 |
| 21-30 | 205 |
| 31-40 | 170 |
| 41-50 | 65 |
| 51-60 | 13 |
| 61-70 | 4 |

## Time in Sweden

| Time | Learners |
|---|---|
| 0 - 2 years | 111 |
| 2 - 4 years | 109 |
| 4 - 6 years | 36 |
| 6 - 8 years | 11 |
| 8 - | 26 |
| N/A | 8 |

**Learner metadata**

Stockholm University

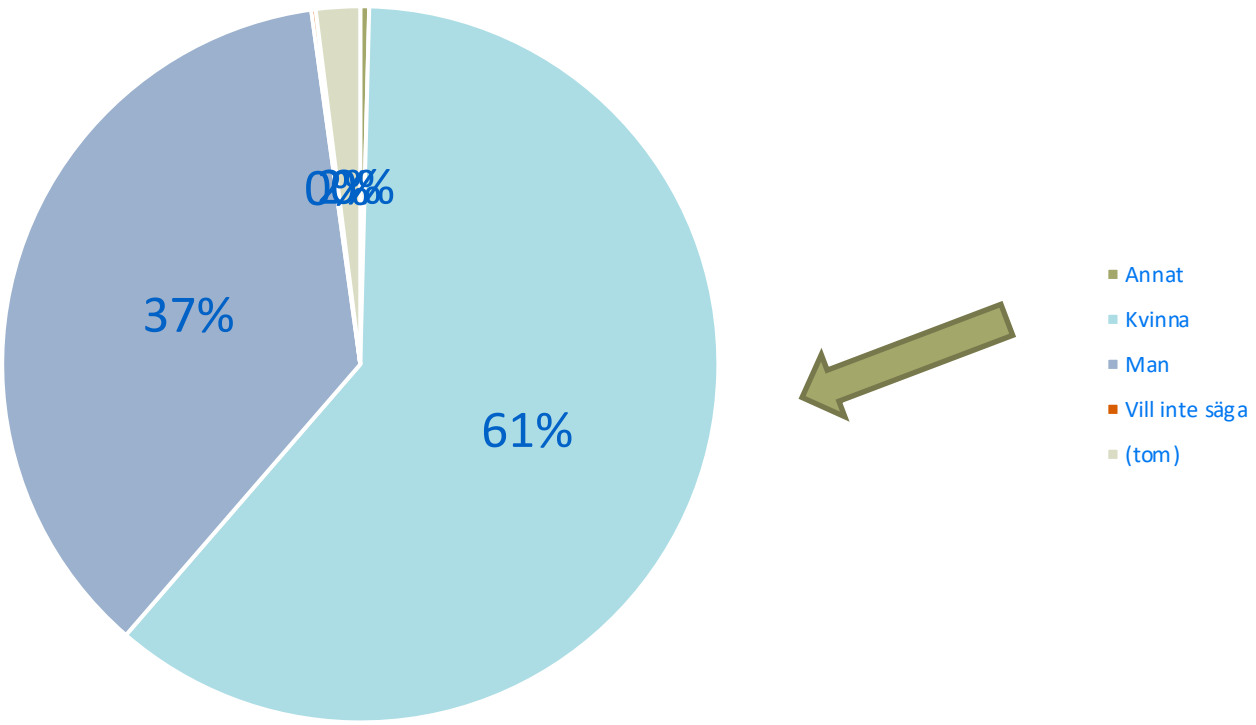
# Proficiency levels & educational background and age

| | Education background |
|---|---|
| **Avancerad** | **168** |
| 1 | 14 |
| 2 | 5 |
| 3 | 70 |
| 4 | 79 |
| **Fortsättning** | **45** |
| 1 | 10 |
| 2 | 13 |
| 3 | 8 |
| 4 | 14 |
| **Nybörjare** | **289** |
| 1 | 10 |
| 2 | 38 |
| 3 | 77 |
| 4 | 164 |
| **Totalsumma** | **502** |

| | Ages |
|---|---|
| **Avancerad** | **168** |
| 16-20 | 33 |
| 21-25 | 65 |
| 26-30 | 24 |
| 31-35 | 19 |
| 36-40 | 13 |
| 41-45 | 6 |
| 46-50 | 3 |
| 51-55 | 1 |
| 56-60 | 1 |
| (tom) | 3 |
| **Fortsättning** | **45** |
| 16-20 | 2 |
| 21-25 | 10 |
| 26-30 | 8 |
| 31-35 | 3 |
| 36-40 | 6 |
| 41-45 | 4 |
| 46-50 | 11 |
| 61-65 | 1 |
| **Nybörjare** | **289** |
| 16-20 | 2 |
| 21-25 | 39 |
| 26-30 | 59 |
| 31-35 | 82 |
| 36-40 | 47 |
| 41-45 | 22 |
| 46-50 | 19 |
| 51-55 | 8 |
| 56-60 | 3 |
| 61-65 | 1 |
| 66-70 | 2 |
| Totalsumma | **502** |

**Learner metadata**

# Gender




## Learner metadata

Stockholm University

# Languages

* 93 learners (21 %) have 2 or more first languages

| First language(s) including bilinguals | | L |
|---|---|---|
| Arabic | appr. 20 % | 94 |
| English | appr. 10 % | 48 |
| Dari + Persian | appr. 10% | 45 |
| Spanish | | 33 |
| Swedish | | 30 |
| Kurdish | | 27 |
| Russian | | 25 |
| Chinese | | 14 |
| German; Turkish | | 13 |
| French | | 12 |
| Greek; Tigrinya | | 11 |
| Somali | | 10 |
| *And more* | | …. |
| ***50-60 languages*** | | |



Stockholm University

# Languages

* 93 learners (21 %) have 2 or more first languages

| First language(s) including bilinguals | L | Best written language* |
|---|---|---|
| Arabic | 94 | 151 |
| English | 48 | 92 |
| Dari + Persian | 45 | 66 |
| Spanish | 33 | 26 |
| Swedish | 30 | 52 |
| Kurdish | 27 | 17 |
| Russian | 25 | 17 |
| Chinese | 14 | 10 |
| German; Turkish | 13 | |
| French | 12 | |
| Greek; Tigrinya | 11 | |
| Somali | 10 | |
| *And more* | … | |
| ***50-60 languages*** | | |