

# NIE22184- Big Data in Education

Presented by Shannamae Yang

Supervised by Asst Prof Farhan Ali & Contributor Vamsi Grandi

## Introduction

We observe the changing trends and ideologies in Education by analyzing language used in over 1 million articles published from 1980 to 2020.

## Method

### Step 1: Separate articles into their respective decades.

Decade 1: 1980-1989

Decade 2: 1990-1999

Decade 3: 2000-2009

Decade 4: 2010-2019

Decade 5: 2020

### Step 2: Make each decades' maps comparable

- Only map terms that appear in all decades
- Only include journals that appear in decade 5

### Step 3: Generate coordinates for each term

Network visualizing software, VosViewer, calculates the coordinates of terms. Term frequency and document frequency is used to calculate the the relevance of terms as well as the distance from each other in a network.

Orientate all maps to have the term “child” at 45 degrees from the x-axis in the top right quadrant so that all decade maps are comparable

### Step 3: Create Global Top 50 terms and edges

**Global Top 50 terms:** The 50 most relevant terms from the all articles.

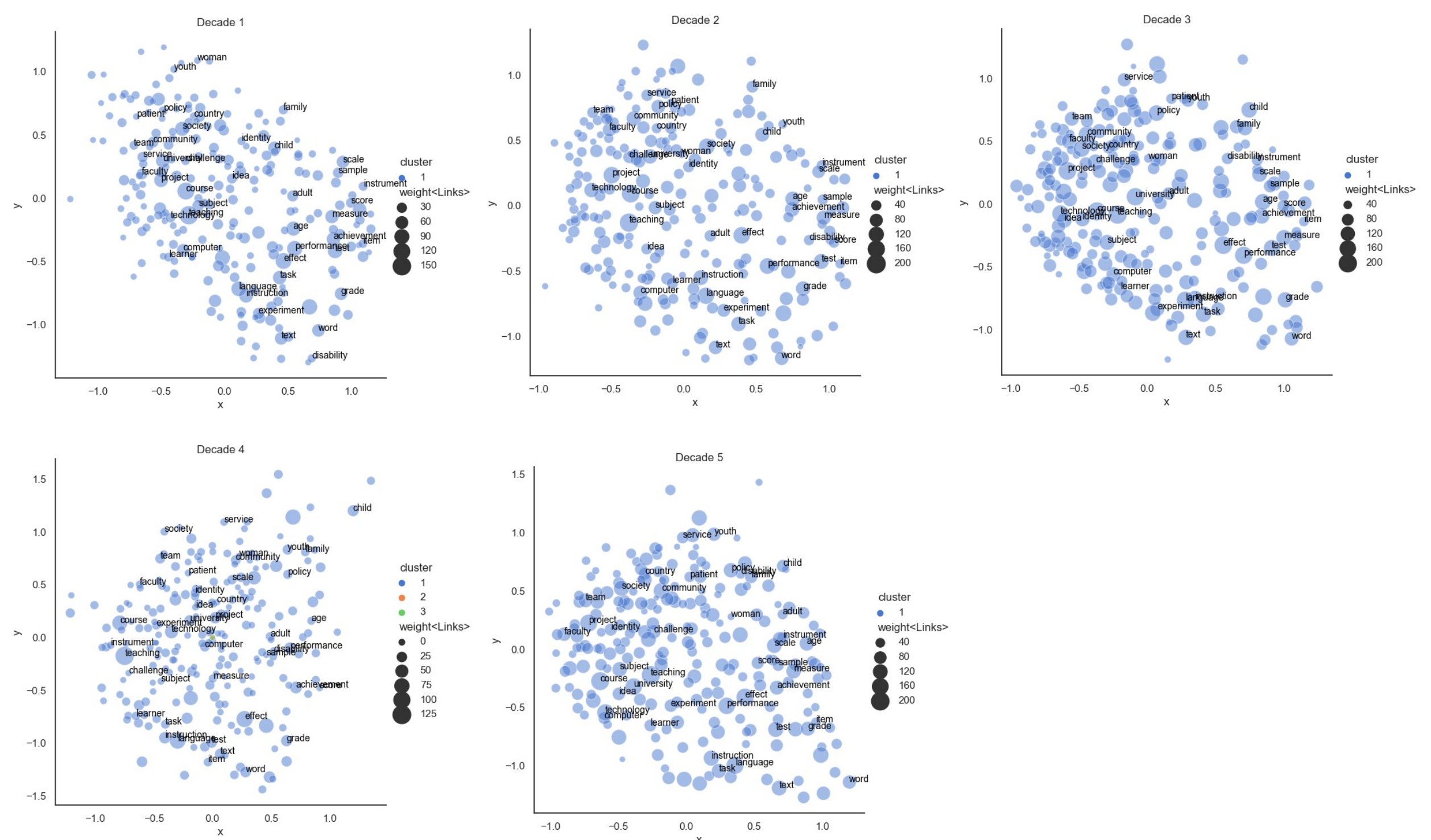
**Edge:** An edge between 2 terms is formed when the 2 terms occur in the same article

**Edge Strength:** Number of articles where the terms occur in together

## Results

### Visualize the Global Top 50 in each Decades

The distance between each node indicates how closely linked they are in that decade.



### Visualize the edges with edge strength >30

An edge indicates that terms appear in over 30 articles and are thus highly relevant in that decade



The term “course” is only linked to teaching and instruction in decade 1. However in all decades afterwards, “course” is linked to more holistic terms; “comprehension”, “engineering”, “feedback”, “disability”, etc

The term “child” and “child care” only becomes highly relevant in decades 2,3 and 4, implying that there is a shift in focus on the well-being of children in these decades.

## Conclusions

Using Big data and Natural Language Processing, we visualize the changes in terminology used in education to deduce how ideologies have changed by observing the following:

- Distance between clusters of words to see how fields become more connected or separated from each other
- Strong edges formed between terms