Data preprocessing

Three datasets were separately collected for Sample 1 (German university students), Sample 2 (German teachers), and Sample 3 (Belgian teachers). On OSF, we share the AMATUS dataset as a partly processed merged dataset. The aim of the preprocessing was (1) to unify and merge the datasets, (2) to exclude ineligible observations according to inclusion/exclusion criteria and quality check, (3) to ensure anonymity, and (4) to make the dataset interpretable and easily reusable. We preprocessed the three datasets and merged them into the AMATUS dataset using R software (version 4.3.2).

The following steps were implemented:

1) Free-entry questions were translated from German/French into English in the original datasets stored as .csv.
2) A custom function called *Preprocessing* was created to perform a series of changes common to all three datasets:
   - recoding dichotomic variables form numeric notation to yes/no notation ('native_speaker', 'sex')
   - standardization of start and end time and calculation of total time in minutes
   - recoding of Belgian math grade into German grade system
   - deleting of uninformative columns (time to complete each questionnaire, free comments at the end of the survey)
   - calculating the number of observations excluded according to each criterium
   - checking arithmetic speeded task entries:
     - finding non-numeric entries and converting it into "-1.0", which is a wrong entry
     - creating a new variable to check if participants skipped a problem (for each arithmetic problem, it assigns 1 if the current field is empty but the following is not; the number of wrongly ordered responses for each participant is summed; whenever the sum is different from 0, the arithmetic score for that participant is excluded)
   - reversing the polarity of items to be reversed coded (see codebook for the specific items)
   - computing total scores for all the questionnaires and the arithmetic task
   - recoding multiple choice quality questions from numeric to string notation
   - reordering columns in the dataset
3) Datasets were merged:
   - in each dataset, columns were renamed
   - the *Preprocessing* function was applied to each dataset
   - for German university students, responses of students that chose "other" in the math load question and then indicated the study program were reassigned to the other three categories (low, medium, high) according to the study program indicated:
     - low: archeology, sport science, social pedagogy, educational sciences, adult education, foreign languages, theology, sinology, culture and history of near east, law, politics
     - medium: geoecology, geoscience, environmental sciences, business administration, economy, psychology, sociology, teacher education in Chemistry
     - high: medical technology, engineering
     - not classified: teacher education, orientation studies
   - the three datasets were pasted together
   - the column 'id_unique' was created, which assigned to each participant a unique numerical ID in sequential order
   - some questions were now recoded from numeric to string: 'breaks', 'device', 'arith_perf_correct_order', 'teacher', 'teacher_stage'
   - some specific errors were corrected in 'teacher_stage': some participants selected the option 'other' and wrote that they are students, so they were assigned to the category 'study'

4) Variables were recategorized for anonymization:
   - We updated teacher_stage: in Sample 2 and 3, three participants chose the option “other” and wrote that their position was either director, retired or teacher who is studying at the same time. They were reassigned to the category “elementary school teachers”
   - we created the following variables:
     - main_focus_mathematics: “yes” was assigned to those who ticked mathematics and “other” and manually entered “basic mathematics”
     - main_focus_literacy: “yes” was assigned to German teachers who had German as main focus and Belgian teachers who had French as main focus
     - main_focus_science: “yes” was assigned to those who ticked science (for Belgian teachers), Sachunterricht (for German teachers) or “other” and manually entered the following subjects: biology, chemistry, geography, physics, science, technology, computer science, handwork/technology
     - main_focus_humanities: “yes” was assigned to those who ticked English, religious pedagogy, art, music, sport, Sachunterricht (for German teachers) or “other” and manually entered the following subjects: economy, everyday culture, everyday culture and health, history, politics, social science, Dutch, pedagogy, psychology
     - teacher_experience: “in-service” was assigned to those who ticked “primary school teacher” in the teacher_stage question, all the others were labeled as “pre-service”
     - age_range: for all participants, a categorical variable with five levels (< 20 years, 20-29 years, 30-39 years, 40-49 years, over 50 years) was created and the age in years for participants in Sample 2 and 3 was deleted
5) Finally, the AMATUS dataset was stored as a .csv file named “AMATUS_dataset.csv” (https://osf.io/gszpb/).