German translation of the Artificial-Social-Agent questionnaire instrument for evaluating human-agent interaction

Underlying Analyses - third translation round

Boleslav Khodakov

06 June, 2023

Contents

1 Introduction

This document presents statistical analyses of correlation and variation between English and German ASA questionnaires for the item level. For the third translation round, the translations of 13 English items are evaluated. There are a total of 52 translations, as all items have multiple alternative translations.

We use the following packages:

```
library(nlme)    # Run multilevel linear models
library(car)     # Package linear regression
library(haven) # Use read_sav fuction
library(dplyr) # Use select function
library(knitr) # Get markdown file
library(tinytex) # Use TeX environment
library(rticles) # Use CTeX documents template
```

```
library(pander) # For pandering tables
panderOptions("table.alignment.default","left")
```

2 Data file

The input data used in the analysis were transformed from the raw data file 'Final_ASA_German_Round_3_anonym.s detailed transformation from raw data to the input data file was explained in the markdown file 'Transformation from raw data to the input data files'.

2.1 File transformed_data_round_3.sav

Human-ASA interaction evaluation data were collected from 30 bilingual participants with German mother tongue who are native German and fluent English speakers. Bilingual participants rated human-ASA interaction on 13 English items and corresponding German translations plus 10 attention control questions. All participants' evaluation data were included as they failed no attention control questions. We removed irrelevant data, e.g., attention control questions, just retaining scores of English items and corresponding German translations, also with 'AgentID'. We did not yet invert reverse-scoring questionnaire items and their English translations. The steps above were conducted and explained in the markdown file 'Transformation from raw data to the input data files', resulting in a single data file 'transformed_data_round_3.sav'. Up to this step, rating scores of 13 English items and corresponding German translations were ready for further analysis.

```
data01 <- data.frame(read_sav("transformed_data_round_3.sav"))
# Select item scores for English and German translation scores
d1 <- select(data01, Q_E_HLA4:Q_DE_AT2_4)
```

3 Analyses results

3.1 Correlation between English and German ASA Questionnaire

We combined the scores of 13 items as well as their corresponding translations in dataframe 'd1'. Then we calculated ICC values for the 13 items. The multilevel model that we fit on the data set is a random intercept model. This model includes a fixed intercept (~1) and participant as a random intercept, indicated by random = ~1|id. Here, 'id' indicates the participant code for 30 bilingual participants whose scores were used to calculate ICC values.

3.1.1 ICC values for 52 items

We calculated ICC values for the 13 items. Items were duplicated leading to a total of 52 (so each of the multiple translations had one match). The multilevel model that we fit on the data set is a random intercept model. This model includes a fixed intercept (~1) and participant as a random intercept,

indicated by random = ~1|id. Here, 'id' indicates the participant code for 30 bilingual participants whose scores were used to calculate ICC values. We calculated ICC as: $\rho_I = \frac{\tau^2}{\tau^2+\sigma^2}$ whereby τ^2 is the variance between participants, and σ^2 is the variance within the score of individual (Finch, Bolin, and Kelley 2019). For the ICC calculation we defined the *getICC* function.

```
getICC <-function(model)
# Function for ICC value calculation using multilevel linear model
{
  vc.model <- VarCorr(model)
  # Estimated variances and correlations between the random-effects terms
  sigma_var <-as.numeric(vc.model[2,1])
  # Variance within the groups
  tau_var <- as.numeric(vc.model[1,1])
  # Variance between the groups
  icc <- tau_var/(tau_var + sigma_var)
  # Calculate ICC value
  return(icc)
}
```

Next, we defined a function to run a multilevel model and obtain the associated ICC value for that model. As input, this function accepts the scores in both languages and the participant ID number. Before the model can be fitted this input data is transformed into a long format. The function returns ICC in value.

```
getLME <-function(s_1,s_2)
# Function for a linear mixed-effects model
{
  id<-rownames(s_2)
  # Row names that represent the ID number of each participant
  score_German<- data.frame(id, s_1, language= 1)
  # Transform German scores from wide format to long format and label as 1
  Score_English<- data.frame(id, s_2, language= 2)
  # Transform English scores from wide format to long format and label as 2
  Score_total <- rbind(score_German, Score_English)
  # Combine German and English scores in the long format
  m0 <- lme(score ~ 1, data = Score_total, random = ~1|id, method = "ML")
  # Linear mixed-effects model with a fixed intercept and
  # a random intercept of participant's ID number
  return(getICC(m0))
}
```

With the *getLME* function defined, the next step is to use this function to calculate the ICC value for

each of the 13 (52, counting duplicated columns for multiple translations) ASA questionnaire items, and in addition, calculate the grand mean of these 52 ICC values. When going to the list of ASAQ items, we use the fact that in the data frame the first 52 columns (incl. duplicates) present the results of the English ASAQ version and the last 52 (unique) columns present the results of the German ASAQ version.

Please note: There is no simple way of displaying only the best (highest-ICC) German items for English items with multiple translations. Below, you will find all translations.

```
l_ICC <- data.frame(ItemID = double(), Item = character(), icc = double())

#Initialize output file for low-ICC combinations
write("low-ICCs for all participants",file="ICC_output.txt",append=TRUE)

# Numbers of columns in d_total
German_column_offset <- ncol(d1) /2

  #Initialize output file for low-ICC combinations
  write("low-ICCs for all participants",file="ICC_output.txt",append=TRUE)

# The value of n is equal to the number of columns divided by 2.
n <- 52
for (i in 1:n)
# Go step by step to 52 items
# whereby i is the ASA questionnaire item number
{

  # Select scores of German version of ASAQ item i
  score_German <- data.frame(score=d1[,i + German_column_offset])

  # Select scores of English version of ASAQ items i
  score_English <- data.frame(score=d1[,i])



  # Calculated ICC and add it to the list of ICC values,
  # with ID number of the ASA questionnaire item
  l_ICC <- rbind(l_ICC, data.frame (i, icc = getLME(score_German, score_English)))

  # Get the current ICC value from l_ICC (which is a table)
  real_ICC <- round(l_ICC[i,2], digits=4)
```

```
  # For values which are under the threshold of 'Good' (0.6)
  if(0.6 > real_ICC){

    # Create an entry in a text file, to have a list of dissatisfactory translations
    output1 <- paste("l_ICC", real_ICC, sep=" ")
    output1 <- paste(output1, colnames(d1[i]), sep=" ")
    output1 <- paste(output1, colnames(d1[i + German_column_offset]), sep=" ")

    # Append a text file (of the user's choice) with the entry
    write(output1,file="ICC_output.txt",append=TRUE)
  }

}

l_ICC$Item = colnames(select(d1,Q_E_HLA4:Q_E_AT2_4)) # Add name code for each item
pander(l_ICC, caption = "ICC values for 52 items")
```

Table 1: ICC values for 52 items

i	icc	Item
1	0.8254	Q_E_HLA4
2	0.6948	Q_E_HLA4_2
3	0.8541	Q_E_HLA4_3
4	0.6369	Q_E_HLA4_4
5	0.3507	Q_E_NA1
6	0.3743	Q_E_NA1_2
7	0.4569	Q_E_NA1_3
8	0.3514	Q_E_NA1_4
9	0.4438	Q_E_UE2
10	0.5012	Q_E_UE2_2
11	0.3736	Q_E_UE2_3
12	0.6245	Q_E_UE2_4
13	0.4058	Q_E_UAL3
14	0.2628	Q_E_UAL3_2
15	0.3941	Q_E_UAL3_3
16	0.4206	Q_E_UAL3_4
17	0.3803	Q_E_UAL5
18	0.6192	Q_E_UAL5_2
19	0.4748	Q_E_UAL5_3

i	icc	Item
20	0.6234	Q_E_UAL5_4
21	0.5777	Q_E_IIS3
22	0.5333	Q_E_IIS3_2
23	0.8343	Q_E_IIS3_3
24	0.541	Q_E_IIS3_4
25	0.5801	Q_E_IIS4
26	0.4606	Q_E_IIS4_2
27	0.3023	Q_E_IIS4_3
28	0.1648	Q_E_IIS4_4
29	0.2725	Q_E_AEI3
30	0.4104	Q_E_AEI3_2
31	0.7619	Q_E_AEI3_3
32	0.6269	Q_E_AEI3_4
33	0.3249	Q_E_AEI5
34	0.5996	Q_E_AEI5_2
35	0.6474	Q_E_AEI5_3
36	0.6091	Q_E_AEI5_4
37	0.63	Q_E_UAI2
38	0.6106	Q_E_UAI2_2
39	0.6477	Q_E_UAI2_3
40	0.5498	Q_E_UAI2_4
41	0.497	Q_E_AEI1
42	0.5642	Q_E_AEI1_2
43	0.2785	Q_E_AEI1_3
44	0.135	Q_E_AEI1_4
45	0.4311	Q_E_UEP2
46	0.3926	Q_E_UEP2_2
47	0.512	Q_E_UEP2_3
48	0.3971	Q_E_UEP2_4
49	0.3467	Q_E_AT2
50	0.3645	Q_E_AT2_2
51	0.5027	Q_E_AT2_3
52	0.2842	Q_E_AT2_4

```
Variable <- c("Grand_mean","SD","Minimum","Maximum")
# Define the names of the statistics
Value <- c(round(mean(l_ICC$icc),digits=4),round(sd(l_ICC$icc),digits=4),
```

```
          round(min(l_ICC$icc),digits=4),round(max(l_ICC$icc),digits=4))
# Calculate the grand mean, standard deviation,
# minimum and maximum values of ICC values of 52 items
description <- cbind(Variable, Value) # Descriptive statistics of ICC values of 52 items

# Print results
pander(description, caption = "Descriptive statistics of ICC values of 13 (52) items")
```

Table 2: Descriptive statistics of ICC values of 13 (52) items

Variable	Value
Grand_mean	0.4896
SD	0.1633
Minimum	0.135
Maximum	0.8541

For the assessment of the correlation between the English and German ASA Questionnaire, we followed Cicchetti's classification of ICC categories (Cicchetti 1994). Then we get the categories of ICC classifications and number of ICC values in classification category. **Please note: There is no simple way of displaying only the best (highest-ICC) German items for English items with multiple translations. Below, you will find the classification of all translations.**

```
Classification <- c("Excellent","Good","Fair","Poor")
ICC_Range <- c("0.75-1.00","0.60-0.74","0.40-0.59","0-0.39")
# Categories of ICC classifications by Cicchetti (1994)
n_item <- length(l_ICC$icc) # Number of ICC values
round_ICC <- round(l_ICC$icc, digits=4) # Round ICC values
Number <- c(length(l_ICC[which(round_ICC>=0.75&round_ICC<=1),]$icc),
            length(l_ICC[which(round_ICC>=0.60&round_ICC<=0.74),]$icc),
            length(l_ICC[which(round_ICC>=0.40&round_ICC<=0.59),]$icc),
            length(l_ICC[which(round_ICC>=0.00&round_ICC<=0.39),]$icc))
# Calculate number of ICC values in classification category
Percentage <- c(round(Number[1]/n_item,digits=4)*100, round(Number[2]/n_item,digits=4)*100,
                round(Number[3]/n_item,digits=4)*100, round(Number[4]/n_item,digits=4)*100)
# Calculate percentage of ICC values in classification category
ICC_category <- cbind(Classification,ICC_Range,Number,Percentage)

# Print results
pander(ICC_category, caption = "Categories of ICC classifications and
```

```
      number of ICC values in classification category for 52 items")
```

Table 3: Categories of ICC classifications and number of ICC values in classification category for 52 items

Classification	ICC_Range	Number	Percentage
Excellent	0.75-1.00	4	7.69
Good	0.60-0.74	11	21.15
Fair	0.40-0.59	18	34.62
Poor	0-0.39	15	28.85

References

Cicchetti, Domenic V. 1994. "Guidelines, Criteria, and Rules of Thumb for Evaluating Normed and Standardized Assessment Instruments in Psychology." *Psychological Assessment* 6 (4): 284. https://doi.org/10.1037/1040-3590.6.4.284.

Finch, W Holmes, Jocelyn E Bolin, and Ken Kelley. 2019. *Multilevel Modeling Using r*. Crc Press.