Appendix 4: Mixed-effect models of match advantages

Planned analysis: Matching was the only one independent variable

1. The model with four random effects: participants, targets, laboratories, and languages.

```
SP_all_random.lmer = lmer(response_time ~ Match + (1|Subject) + (1|Target) + (1|PSA_ID) + (1|Language), data = SP_V_lme_data) ## build mixed-effec

tab_model(SP_all_random.lmer, title = "Coefficients")
```

Coefficients

	response_time		
Predictors	Estimates	CI	p
(Intercept)	907.20	742.26 – 1072.13	**<0.001**
Match [MISMATCHING]	-14.42	-65.11 – 36.27	0.577
Random Effects			
σ^2	10100135.94		
$\tau_{00\ Subject}$	0.00		
$\tau_{00\ PSA\_ID}$	122977.83		
$\tau_{00\ Target}$	0.00		
$\tau_{00\ Language}$	50155.84		
$N_{Subject}$	2692		
N_{Target}	48		
N_{PSA_ID}	50		
$N_{Language}$	18		
Observations	60404		
Marginal R^2 / Conditional R^2	0.000 / NA		

2. The model excluded the targets and languages from the random effect structure. The final report decided `SP_reduced_random.lmer` the best fitted model.

```
SP_reduced_random.lmer = lmerTest::lmer(response_time ~ Match + (1|Subject) + (1|PSA_ID), data = SP_V_lme_data)
SP_slope_nocor_reduced_random.lmer = lmer(response_time ~ Match + (1|Subject) + (Match||PSA_ID), data = SP_V_lme_data)
SP_slope_cor_reduced_random.lmer = lmer(response_time ~ Match + (1|Subject) + (Match|PSA_ID), data = SP_V_lme_data)

tab_model(SP_reduced_random.lmer)
```

	response_time		
Predictors	Estimates	CI	p
(Intercept)	905.23	788.92 – 1021.53	**<0.001**
Match [MISMATCHING]	-14.43	-65.12 – 36.26	0.577
Random Effects			
σ^2	10100120.51		
$\tau_{00\ Subject}$	0.00		
$\tau_{00\ PSA\_ID}$	155340.45		
$N_{Subject}$	2692		
N_{PSA_ID}	50		
Observations	60404		
Marginal R^2 / Conditional R^2	0.000 / NA		

```
tab_model(SP_slope_nocor_reduced_random.lmer)
```

	response_time		
Predictors	Estimates	CI	p
(Intercept)	904.35	786.48 – 1022.21	**<0.001**
Match [MISMATCHING]	-12.75	-63.54 – 38.05	0.623
$N_{Subject}$	2692		
N_{PSA_ID}	50		
Observations	60404		

```
tab_model(SP_slope_cor_reduced_random.lmer)
```

	response_time		
Predictors	Estimates	CI	p
(Intercept)	904.35	786.52 – 1022.17	**<0.001**
Match [MISMATCHING]	-12.75	-63.55 – 38.04	0.623
Random Effects			
σ^2	10100085.72		
$\tau_{00\ Subject}$	0.00		
$\tau_{00\ PSA\_ID}$	159848.14		
$\tau_{11\ PSA\_ID.MatchMISMATCHING}$	133.25		
$\rho_{01\ PSA\_ID}$	-1.00		
$N_{Subject}$	2692		
N_{PSA_ID}	50		
Observations	60404		
Marginal R^2 / Conditional R^2	0.000 / NA		

Models included data source

We evaluated the interaction of match advantage and data collection sources in three models. Based on the recommended practices(Barr et al., 2013; Brauer & Curtin, 2018), the models used the optimizer `bobyqa`. The final report decided `source_cor.lmer` the best fitted model.

```
## standardized
SP_V_lme_data$r_Source = if_else(SP_V_lme_data$Source == "Site",1,0)

source_zero_slope_nocor.lmer =lmer(response_time ~ Match*r_Source + (1|Subject) + (1|PSA_ID) + (1|Language),
  control = lmerControl(optimizer = "bobyqa",
                        optCtrl = list(maxfun = 1e6)), data = SP_V_lme_data)


source_nocor.lmer =lmer(response_time ~ Match*r_Source + (1|Subject) + (r_Source||PSA_ID) + (r_Source||Language),
  control = lmerControl(optimizer = "bobyqa",
                        optCtrl = list(maxfun = 1e6)),
   data = SP_V_lme_data)

source_cor.lmer =lmer(response_time ~ Match*r_Source + (1|Subject) + (r_Source|PSA_ID) + (r_Source|Language),
  control = lmerControl(optimizer = "bobyqa",
                        optCtrl = list(maxfun = 1e6)),
                        data = SP_V_lme_data)

tab_model(source_zero_slope_nocor.lmer)
```

	response_time		
Predictors	Estimates	CI	p
(Intercept)	1360.41	1236.16 – 1484.67	**<0.001**
Match [MISMATCHING]	-39.57	-117.06 – 37.91	0.317
r_Source	-685.06	-793.42 – -576.71	**<0.001**
Match [MISMATCHING] * r_Source	43.94	-58.50 – 146.39	0.400
Random Effects			
σ^2	10099089.93		
$\tau_{00\ Subject}$	0.00		
$\tau_{00\ PSA\_ID}$	13643.18		
$\tau_{00\ Language}$	31903.35		
$N_{Subject}$	2692		
N_{PSA_ID}	50		
$N_{Language}$	18		
Observations	60404		
Marginal R^2 / Conditional R^2	0.011 / NA		

```
tab_model(source_nocor.lmer)
```

	response_time		
Predictors	Estimates	CI	p
(Intercept)	1360.41	1236.16 – 1484.67	**<0.001**
Match [MISMATCHING]	-39.57	-117.06 – 37.91	0.317
r_Source	-685.06	-793.42 – -576.71	**<0.001**
Match [MISMATCHING] * r_Source	43.94	-58.50 – 146.39	0.400
Random Effects			
σ^2	10099089.92		
$\tau_{00\ Subject}$	0.00		
$\tau_{00\ PSA\_ID}$	13643.25		
$\tau_{00\ Language}$	31903.23		
$\tau_{11\ PSA\_ID.r\_Source}$	0.00		
$\tau_{11\ Language.r\_Source}$	0.00		
ρ_{01}			
ρ_{01}			
$N_{Subject}$	2692		
N_{PSA_ID}	50		
$N_{Language}$	18		
Observations	60404		
Marginal R^2 / Conditional R^2	0.011 / NA		

```
tab_model(source_cor.lmer)
```

	response_time		
Predictors	Estimates	CI	p
(Intercept)	1374.45	1145.16 – 1603.75	**<0.001**
Match [MISMATCHING]	-39.70	-117.17 – 37.77	0.315
r_Source	-744.57	-978.73 – -510.42	**<0.001**
Match [MISMATCHING] * r_Source	44.01	-58.41 – 146.43	0.400
Random Effects			
σ^2	10094277.36		
$\tau_{00\ Subject}$	0.00		
$\tau_{00\ PSA\_ID}$	46683.47		
$\tau_{00\ Language}$	79313.45		
$\tau_{11\ PSA\_ID.r\_Source}$	47754.95		
$\tau_{11\ Language.r\_Source}$	79299.01		
$\rho_{01\ PSA\_ID}$	-1.00		
$\rho_{01\ Language}$	-1.00		

$N_{Subject}$	2692
N_{PSA_ID}	50
$N_{Language}$	18
Observations	60404
Marginal R^2 / Conditional R^2	0.013 / NA

Models included languages

We analyzed the interactions by the data sources separately.

on site data

We evaluated the interaction of match advantage and languagess in three models. Based on the recommended practices(Barr et al., 2013; Brauer & Curtin, 2018), the models used the optimizer `bobyqa`. The final report decided `lang_cor.lmer` the best fitted model.

```
## Check sample size of a language by site data
site_excluded_lang <- subset(SP_V_lme_data, Source=="Site") %>%
  group_by(Language, Subject) %>%
  summarise(N_trials = n()) %>%
  group_by(Language) %>%
  summarise(N = n()) %>%
  filter(N < 25) %>%
  pull(Language)

## Allocate the site data
SP_V_lang_lme_data = subset(SP_V_lme_data, Source=="Site" & !(Language %in% site_excluded_lang))

## Run the mixed effect model by site data
lang_cor.lmer =lmer(response_time ~ Language*Match + (1|Subject),
  control = lmerControl(optimizer = "bobyqa",
                        optCtrl = list(maxfun = 1e6)),
  data = SP_V_lang_lme_data)

lang_slope_nocor.lmer =lmer(response_time ~ Language*Match + (Match||Subject),
  control = lmerControl(optimizer = "bobyqa",
                        optCtrl = list(maxfun = 1e6)), data = SP_V_lang_lme_data)

lang_slope_cor.lmer =lmer(response_time ~ Language*Match + (Match|Subject),
  control = lmerControl(optimizer = "bobyqa",
                        optCtrl = list(maxfun = 1e6)), data = SP_V_lang_lme_data)


tab_model(lang_cor.lmer)
```

	response_time		
Predictors	Estimates	CI	p
(Intercept)	603.90	594.69 – 613.11	**<0.001**
Language [German]	-30.29	-55.28 – -5.30	**0.018**
Language [Greek]	123.21	97.90 – 148.51	**<0.001**
Language [Hebrew]	-15.66	-36.91 – 5.58	0.148
Language [Hindi]	66.91	39.13 – 94.69	**<0.001**
Language [Hungarian]	39.07	16.78 – 61.35	**0.001**
Language [Norwegian]	5.58	-17.18 – 28.33	0.631
Language [Polish]	0.98	-33.13 – 35.10	0.955
Language [Simplified Chinese]	59.20	31.69 – 86.70	**<0.001**
Language [Slovak]	23.06	1.29 – 44.82	**0.038**
Language [Spanish]	83.98	61.39 – 106.56	**<0.001**
Language [Thai]	39.78	5.19 – 74.38	**0.024**
Language [Traditional Chinese]	54.33	29.81 – 78.84	**<0.001**
Language [Turkish]	59.02	39.60 – 78.44	**<0.001**
Match [MISMATCHING]	5.73	-1.32 – 12.77	0.111
Language [German] * Match [MISMATCHING]	9.93	-9.10 – 28.96	0.306
Language [Greek] * Match [MISMATCHING]	21.72	2.12 – 41.31	**0.030**
Language [Hebrew] * Match [MISMATCHING]	1.95	-14.17 – 18.08	0.812
Language [Hindi] * Match [MISMATCHING]	-14.07	-35.75 – 7.61	0.203
Language [Hungarian] * Match [MISMATCHING]	-14.08	-31.05 – 2.90	0.104
Language [Norwegian] * Match [MISMATCHING]	0.66	-16.63 – 17.94	0.941
Language [Polish] * Match [MISMATCHING]	-10.98	-37.00 – 15.03	0.408
Language [Simplified Chinese] * Match [MISMATCHING]	-5.93	-27.13 – 15.28	0.584
Language [Slovak] * Match [MISMATCHING]	-1.65	-18.20 – 14.89	0.845
Language [Spanish] * Match [MISMATCHING]	-6.98	-24.34 – 10.38	0.430
Language [Thai] * Match [MISMATCHING]	-5.11	-31.96 – 21.73	0.709
Language [Traditional Chinese] * Match [MISMATCHING]	5.80	-12.96 – 24.56	0.544

Language [Turkish] * Match [MISMATCHING]	-9.95	-24.76 – 4.85	0.188
Random Effects			
σ^2	34460.57		
$\tau_{00\ Subject}$	7389.48		
ICC	0.18		
$N_{\ Subject}$	1524		
Observations	34441		
Marginal R^2 / Conditional R^2	0.034 / 0.204		

```
tab_model(lang_slope_nocor.lmer)
```

	response_time		
Predictors	Estimates	CI	p
(Intercept)	603.89	594.75 – 613.04	**<0.001**
Language [German]	-30.28	-55.09 – -5.47	**0.017**
Language [Greek]	123.20	98.09 – 148.32	**<0.001**
Language [Hebrew]	-15.65	-36.74 – 5.43	0.146
Language [Hindi]	66.89	39.31 – 94.46	**<0.001**
Language [Hungarian]	39.08	16.95 – 61.20	**0.001**
Language [Norwegian]	5.59	-17.00 – 28.18	0.628
Language [Polish]	0.98	-32.88 – 34.85	0.955
Language [Simplified Chinese]	59.21	31.90 – 86.51	**<0.001**
Language [Slovak]	23.06	1.46 – 44.66	**0.036**
Language [Spanish]	83.95	61.53 – 106.37	**<0.001**
Language [Thai]	39.76	5.42 – 74.10	**0.023**
Language [Traditional Chinese]	54.33	30.00 – 78.67	**<0.001**
Language [Turkish]	59.02	39.74 – 78.29	**<0.001**
Match [MISMATCHING]	5.73	-1.36 – 12.83	0.113
Language [German] * Match [MISMATCHING]	9.91	-9.24 – 29.07	0.310
Language [Greek] * Match [MISMATCHING]	21.75	2.03 – 41.47	**0.031**
Language [Hebrew] * Match [MISMATCHING]	1.94	-14.29 – 18.17	0.815
Language [Hindi] * Match [MISMATCHING]	-14.08	-35.90 – 7.74	0.206
Language [Hungarian] * Match [MISMATCHING]	-14.09	-31.17 – 3.00	0.106
Language [Norwegian] * Match [MISMATCHING]	0.65	-16.75 – 18.05	0.942
Language [Polish] * Match [MISMATCHING]	-10.98	-37.17 – 15.21	0.411
Language [Simplified Chinese] * Match [MISMATCHING]	-5.92	-27.26 – 15.42	0.587
Language [Slovak] * Match [MISMATCHING]	-1.64	-18.30 – 15.01	0.847
Language [Spanish] * Match [MISMATCHING]	-6.96	-24.43 – 10.51	0.435
Language [Thai] * Match [MISMATCHING]	-5.08	-32.10 – 21.93	0.712
Language [Traditional Chinese] * Match [MISMATCHING]	5.81	-13.08 – 24.69	0.547
Language [Turkish] * Match [MISMATCHING]	-9.94	-24.85 – 4.97	0.191
$N_{\ Subject}$	1524		
Observations	34441		

```
tab_model(lang_slope_cor.lmer)
```

	response_time		
Predictors	Estimates	CI	p
(Intercept)	603.89	594.75 – 613.04	**<0.001**
Language [German]	-30.28	-55.09 – -5.47	**0.017**
Language [Greek]	123.20	98.09 – 148.32	**<0.001**
Language [Hebrew]	-15.65	-36.74 – 5.43	0.146
Language [Hindi]	66.89	39.31 – 94.46	**<0.001**
Language [Hungarian]	39.08	16.95 – 61.20	**0.001**
Language [Norwegian]	5.59	-17.00 – 28.18	0.628
Language [Polish]	0.98	-32.88 – 34.85	0.955
Language [Simplified Chinese]	59.21	31.90 – 86.51	**<0.001**
Language [Slovak]	23.06	1.46 – 44.66	**0.036**
Language [Spanish]	83.95	61.53 – 106.37	**<0.001**
Language [Thai]	39.76	5.42 – 74.10	**0.023**
Language [Traditional Chinese]	54.33	30.00 – 78.67	**<0.001**

Language [Turkish]	59.02	39.74 – 78.29	**<0.001**
Match [MISMATCHING]	5.73	-1.36 – 12.83	0.113
Language [German] * Match [MISMATCHING]	9.91	-9.24 – 29.07	0.310
Language [Greek] * Match [MISMATCHING]	21.75	2.03 – 41.47	**0.031**
Language [Hebrew] * Match [MISMATCHING]	1.94	-14.29 – 18.17	0.815
Language [Hindi] * Match [MISMATCHING]	-14.08	-35.90 – 7.74	0.206
Language [Hungarian] * Match [MISMATCHING]	-14.09	-31.17 – 3.00	0.106
Language [Norwegian] * Match [MISMATCHING]	0.65	-16.75 – 18.05	0.942
Language [Polish] * Match [MISMATCHING]	-10.98	-37.17 – 15.21	0.411
Language [Simplified Chinese] * Match [MISMATCHING]	-5.92	-27.26 – 15.42	0.587
Language [Slovak] * Match [MISMATCHING]	-1.64	-18.30 – 15.01	0.847
Language [Spanish] * Match [MISMATCHING]	-6.96	-24.43 – 10.51	0.435
Language [Thai] * Match [MISMATCHING]	-5.08	-32.10 – 21.93	0.712
Language [Traditional Chinese] * Match [MISMATCHING]	5.81	-13.08 – 24.69	0.547
Language [Turkish] * Match [MISMATCHING]	-9.94	-24.85 – 4.97	0.191
Random Effects			
σ^2	34438.84		
$\tau_{00\ Subject}$	7238.67		
$\tau_{11\ Subject.MatchMISMATCHING}$	83.90		
$\rho_{01\ Subject}$	0.17		
ICC	0.18		
$N_{Subject}$	1524		
Observations	34441		
Marginal R^2 / Conditional R^2	0.034 / 0.205		

web-based data

We evaluated the interaction of match advantage and languagess in three models. Based on the recommended practices(Barr et al., 2013; Brauer & Curtin, 2018), the models used the optimizer `bobyqa`. The final report decided `osweb_cor.lmer` the best fitted model.

```
## Check sample size of a language by osweb data
osweb_excluded_lang <- subset(SP_V_lme_data, Source=="Internet") %>%
  group_by(Language, Subject) %>%
  summarise(N_trials = n()) %>%
  group_by(Language) %>%
  summarise(N = n()) %>%
  filter(N < 25) %>%
  pull(Language)  ## Exclude the languages less than 25 participants

## Allocate the osweb data
SP_V_osweb_lme_data = subset(SP_V_lme_data, Source=="Internet" & !(Language %in% osweb_excluded_lang))


## Run the mixed effect model by site data
osweb_cor.lmer = lmerTest::lmer(response_time ~ Language*Match + (1|Subject),
  control = lmerControl(optimizer = "bobyqa",
                        optCtrl = list(maxfun = 1e6)), # Increase maximum number of iterations to facilitate model convergence ,
        data = SP_V_osweb_lme_data)

osweb_slope_nocor.lmer =lmer(response_time ~ Language*Match + (Match||Subject),
  control = lmerControl(optimizer = "bobyqa",
        optCtrl = list(maxfun = 1e6)), data = SP_V_osweb_lme_data)

osweb_slope_cor.lmer =lmer(response_time ~ Language*Match + (Match|Subject),
  control = lmerControl(optimizer = "bobyqa",
        optCtrl = list(maxfun = 1e6)), data = SP_V_osweb_lme_data)


tab_model(osweb_cor.lmer)
```

	response_time		
Predictors	Estimates	CI	p
(Intercept)	908.78	553.58 – 1263.99	**<0.001**
Language [Brazilian Portuguese]	735.18	197.77 – 1272.59	**0.007**
Language [English]	397.20	24.14 – 770.25	**0.037**
Language [German]	875.24	440.05 – 1310.43	**<0.001**
Language [Portuguese]	346.34	-226.17 – 918.85	0.236
Language [Serbian]	945.80	511.11 – 1380.49	**<0.001**
Language [Traditional Chinese]	64.78	-536.87 – 666.42	0.833
Language [Turkish]	510.13	-8.57 – 1028.83	0.054
Match [MISMATCHING]	-3.92	-504.22 – 496.37	0.988

Language [Brazilian Portuguese] * Match [MISMATCHING]	-140.28	-898.94 – 618.38	0.717
Language [English] * Match [MISMATCHING]	-57.33	-583.00 – 468.34	0.831
Language [German] * Match [MISMATCHING]	-272.91	-886.48 – 340.66	0.383
Language [Portuguese] * Match [MISMATCHING]	68.91	-737.32 – 875.14	0.867
Language [Serbian] * Match [MISMATCHING]	277.47	-336.10 – 891.04	0.375
Language [Traditional Chinese] * Match [MISMATCHING]	-26.12	-882.09 – 829.84	0.952
Language [Turkish] * Match [MISMATCHING]	-36.82	-766.74 – 693.10	0.921
Random Effects			
σ^2	23878104.68		
$\tau_{00\ Subject}$	0.00		
$N_{\ Subject}$	1147		
Observations	25480		
Marginal R^2 / Conditional R^2	0.003 / NA		

```
tab_model(osweb_slope_nocor.lmer)
```

	response_time		
Predictors	Estimates	CI	p
(Intercept)	908.89	553.03 – 1264.76	**<0.001**
Language [Brazilian Portuguese]	735.08	196.58 – 1273.58	**0.007**
Language [English]	397.07	23.31 – 770.83	**0.037**
Language [German]	875.10	439.06 – 1311.15	**<0.001**
Language [Portuguese]	346.24	-227.45 – 919.93	0.237
Language [Serbian]	945.91	510.36 – 1381.45	**<0.001**
Language [Traditional Chinese]	64.65	-538.23 – 667.54	0.834
Language [Turkish]	510.01	-9.72 – 1029.74	0.054
Match [MISMATCHING]	-4.15	-506.95 – 498.65	0.987
Language [Brazilian Portuguese] * Match [MISMATCHING]	-140.02	-902.82 – 622.77	0.719
Language [English] * Match [MISMATCHING]	-57.10	-585.44 – 471.24	0.832
Language [German] * Match [MISMATCHING]	-272.66	-889.48 – 344.15	0.386
Language [Portuguese] * Match [MISMATCHING]	69.14	-741.55 – 879.84	0.867
Language [Serbian] * Match [MISMATCHING]	277.40	-339.40 – 894.20	0.378
Language [Traditional Chinese] * Match [MISMATCHING]	-25.88	-886.53 – 834.77	0.953
Language [Turkish] * Match [MISMATCHING]	-36.57	-770.40 – 697.26	0.922
$N_{\ Subject}$	1147		
Observations	25480		

```
tab_model(osweb_slope_cor.lmer)
```

	response_time		
Predictors	Estimates	CI	p
(Intercept)	908.89	553.03 – 1264.76	**<0.001**
Language [Brazilian Portuguese]	735.08	196.58 – 1273.58	**0.007**
Language [English]	397.07	23.31 – 770.83	**0.037**
Language [German]	875.10	439.06 – 1311.15	**<0.001**
Language [Portuguese]	346.24	-227.45 – 919.93	0.237
Language [Serbian]	945.91	510.36 – 1381.45	**<0.001**
Language [Traditional Chinese]	64.65	-538.23 – 667.54	0.834
Language [Turkish]	510.01	-9.72 – 1029.74	0.054
Match [MISMATCHING]	-4.15	-506.95 – 498.65	0.987
Language [Brazilian Portuguese] * Match [MISMATCHING]	-140.02	-902.82 – 622.77	0.719
Language [English] * Match [MISMATCHING]	-57.10	-585.44 – 471.24	0.832
Language [German] * Match [MISMATCHING]	-272.66	-889.48 – 344.15	0.386
Language [Portuguese] * Match [MISMATCHING]	69.14	-741.56 – 879.84	0.867
Language [Serbian] * Match [MISMATCHING]	277.40	-339.40 – 894.20	0.378

Language [Traditional Chinese] * Match [MISMATCHING]	-25.88	-886.53 – 834.77	0.953
Language [Turkish] * Match [MISMATCHING]	-36.57	-770.40 – 697.26	0.922
Random Effects			
σ^2	23865260.44		
$\tau_{00\ Subject}$	10220.10		
$\tau_{11\ Subject.MatchMISMATCHING}$	51127.34		
$\rho_{01\ Subject}$	-1.00		
$N_{\ Subject}$	1147		
Observations	25480		
Marginal R^2 / Conditional R^2	0.003 / NA		