Going global: 39 language versions of the BFI-2-XS

Beatrice Rammstedt,

Lena Roemer,

Dorothée Behr,

Matthias Bluemke,

Clemens Lechner,

Steve Dept,

Laura Wäyrynen,

Chris Soto,

Oliver P. John

OECD

DIRECTORATE FOR EDUCATION AND SKILLS

Going Global: 39 Language Versions of the BFI-2-XS

OECD Education Working Paper No. 309

By Beatrice Rammstedt (GESIS and University of Mannheim), Lena Roemer (GESIS), Dorothée Behr (GESIS), Matthias Bluemke (GESIS), Clemens Lechner (GESIS), Steve Dept (cApStAn), Laura Wäyrynen (cApStAn), Chris Soto (Colby College), and Oliver P. John (University of California, Berkeley)

This working paper has been authorised by Andreas Schleicher, Director of the Directorate for Education and Skills, OECD.

Beatrice Rammstedt, beatrice.rammstedt@gesis.org

JT03539378

<div align="center">OECD EDUCATION WORKING PAPERS SERIES</div>

<div align="center">--

www.oecd.org/edu/workingpapers

--</div>

Abstract

In the Second Cycle of the Survey of Adult Skills (PIAAC), whose results will be released at the end of 2024, the Big Five personality traits were assessed using the BFI-2-XS, the 15-item extra-short form of the Big Five Inventory-2 (BFI-2). For this purpose, the instrument was translated into 24 languages and adapted to 29 countries, resulting in 39 language versions. This translation and adaptation process followed state-of-the-art procedures to generate language versions of the BFI-2-XS that are maximally comparable across countries and regions. In the present paper, we describe this general translation procedure from a methodological point of view. We also document each resulting language version and report in detail the decisions taken during the translation process and the adaptations made to preexisting national versions of the BFI-2-XS. Our aim is to share with researchers the resulting BFI-2-XS language versions developed with high quality standards to allow maximal cross-cultural comparability. Our intention in so doing is to enable their wider usage beyond PIAAC.

Table of contents

FIGURES

TABLES

1. Introduction

In recent decades, the Big Five personality traits have gained wide acceptance as a framework to parsimoniously describe personality on a global level (John, Naumann and Soto, 2008[1]; McCrae and Costa, 2008[2]). This has led to broad interest in their assessment, even in fields beyond core personality research, such as sociology, economics, and public health. Numerous cross-cultural surveys, such as the World Values Survey, or WVS (Ludeke and Larsen, 2017[3]) and the Survey of Health, Ageing, and Retirement in Europe, or SHARE (Levinsky, Litwin and Lechner, 2019[4]), now include measures of the broad personality dimensions in order to compare cultural differences in the impact of these dimensions on the constructs of interest.

Large-scale survey studies, such as the WVS or SHARE, are characterised by the fact that (a) they focus on topics other than personality, and (b) they aim to assess the Big Five domains as outcomes or correlates at the population level rather than for individual diagnostic purposes (e.g. assessment). To meet the demands of these surveys, there is a need for ultra-efficient measures—of the Big Five as well as other constructs of interest— whose psychometric properties allow group comparisons.

Several ultra-short instruments assessing the Big Five have been developed in recent decades. One of the most prominent and widely used instruments in large-scale assessments is the 10-item Big Five Inventory, or BFI-10 (Rammstedt and John, 2007[5])[1] 2023), an ultrashort version of the established 44-item Big Five Inventory, known as BFI-44 (John, Donahue and Kentle, 1991[6]; John, Naumann and Soto, 2008[1]).

Such ultra-short measures assess the global Big Five domains, but they cannot reflect their heterogeneity (i.e. the fact that each broad domain subsumes several more-specific facet traits). As researchers have become more and more interested in assessing not only the global domains but also the facets of the Big Five, instruments reflecting this hierarchical structure, such as the 60-item BFI-2 (Soto and John, 2017[7]), have been developed. The BFI-2 distinguishes between three central facets in each Big Five domain (e.g. Sociability, Assertiveness, and Energy Level within the Extraversion domain). To allow the assessment of both global and facet levels in research contexts with time limitations, Soto and John (2017[8]) developed and validated two abbreviated forms of the BFI-2: (a) the 30-item BFI-2-S and (b) the BFI-2-XS, comprising a 15-item subset of the 30 BFI-2-S items. By including one item from each of the three facets defining each Big Five domain, the 3-item domain scales of the BFI-2-XS cover the full breadth of the Big Five dimensions as defined in the original BFI-2. A more technical advantage of the BFI-2-XS is that each of its domains can be modelled as a latent variable, because three items are enough for a latent variable to be (just) identified in confirmatory factor analysis.

2. The Survey of Adult Skills (PIAAC)

The Survey of Adult Skills (PIAAC) is an established international comparative large-scale study programme initiated and run by the Organisation for Economic Co-operation and Development (OECD). It assesses key cognitive skills, such as literacy and numeracy, in the adult population. As PIAAC aims to draw conclusions about human capital and its usage in the various participating countries, it is crucial that all measures used be comparable across countries. To achieve this, PIAAC ensures compliance with very high

---

[1] For an overview of its usage and psychometric performance, see Rammstedt, Roemer and Lechner (2023[21]).

methodological standards for cross-cultural translations and adaptations. It also includes a large-scale Field Trial conducted in all participating countries to test the instruments and the study design before the Main Study is fielded.[2]

The OECD commissioned an international consortium to plan the design of PIAAC and to monitor its implementation in the participating countries. The PIAAC Consortium member with primary responsibility for the translation and cultural adaptation of the survey instruments is the language service provider cApStAn. Within each country, national centres were established to implement PIAAC according to a detailed set of methodological standards to ensure comparability across countries and the overall quality of the resulting data.

In its current second cycle, in addition to cognitive skills, PIAAC also assessed some basic non-cognitive skills as additional background variables. For this purpose, it was decided to include a measure assessing the Big Five dimensions of personality. To investigate which Big Five instrument best fit its needs, the OECD convened an expert panel and conducted comprehensive national and international pilot studies in which the full BFI-2 was assessed. Results of these pilots suggested that the variance explained in numerous outcomes of interest to PIAAC was highly comparable for the 60-item BFI-2, the 30-item BFI-S, and the 15-item BFI-XS (Rammstedt, Lechner and Danner, 2024[9]). Due to time constraints in the administration of the background questionnaire, the OECD initially decided to assess only the two dimensions of Conscientiousness and Openness with six items each from the BFI-2-S in the PIAAC Field Trial. 24 countries decided nevertheless to administer the full 30-items BFI-2-S in the Field Trial. For the Main Study, the decision was taken to cover all dimensions, but to use the shorter BFI-2-XS instrument to save time. Participating countries were given the option to administer the additional items from the BFI-2-S if they wished. Out of the 28 countries that opted to participate in the Big Five assessment in the Main Study, 12 decided to test the full 30-item BFI-2-S in the Main Study. In sum, the instruments were translated into 23 languages and adapted to 29 countries, resulting in 39 language versions.

3. Aims of the present study

The aims of the present study are to make these language versions of the BFI-2-S and BFI-2-XS available to a wider research audience and to present the general state-of-the-art translation and adaptation procedure followed in PIAAC. In the Results section, we will describe the major decisions taken and adaptations made in the multi-step development of the final BFI-2-XS fielded in the Main Study of PIAAC. In doing so, we will adopt a national or, where necessary, a language-by-country perspective. Details of the adaptation steps by country and item will be documented in an online appendix. Finally, and most importantly, we will share the resulting final translated BFI-2-XS versions (and BFI-2-S versions if available) to enable their use in the broader scientific community. Although OECD guidelines preclude the publication of empirical results from the PIAAC Field Trial, widely sharing these language versions of the BFI-2-S and BFI-2-XS, as well as the procedures used to develop and validate them, will facilitate both within-country and international research on personality traits and non-cognitive skills in a broad range of languages and cultures around the world.

---

[2] Due to fieldwork constraints during the COVID-19 pandemic, a small subset of countries employed only simulated data to study the design and emulate the processes.

4. Method

Thirty-one countries participated in the current second cycle of PIAAC (referred to in what follows as PIAAC 2023), which was fielded in 2022–23. Data and results are due to be published in 2024. The PIAAC 2023 survey comprised a detailed background questionnaire administered in the form of a personal interview, followed by an in-depth assessment of cognitive skills completed by the respondents in the presence of the interviewer.

The non-cognitive skills module, which comprised the BFI-2-XS, was an internationally optional element of the PIAAC 2023 background questionnaire. Thus, countries could decide whether or not to include it. Of the 31 participating countries, only three – Korea, Japan and the United States – opted not to do so. Korea, however, administered the BFI-2-S instruments in the Field Trial and even implemented changes after the Field Trial. The Korean adaptation of the BFI-2-S thus fully followed process described below and are therefore included in the present documentation.

5. Countries and language regions

The 29 countries that developed the non-cognitive skills module for PIAAC 2023 are listed in Table 1. In eight of these countries, the module was administered in multiple languages (e.g. Spain, where regional languages such as Basque, Catalan and Galician are used in addition to Spanish). For the different language groups within these countries, separate instruments were prepared. Some countries (e.g. Germany, Austria, and Switzerland) share an official language. However, as the spoken language differs slightly between these countries, different language versions were prepared, albeit from a common starting point. This resulted in a total of 39 different language versions of the non-cognitive skills module in PIAAC 2023.

Table 1. Translation and adaptation steps and decisions for the 39 language versions of the BFI-2-(X)S

Language	Country	Adaptations of the BFI-2-S for PIAAC Field Trial						Adaptations of the BFI-2-XS for PIAAC Main Study				Comments
		BFI-2 version in FT [1]	Pre-existing version? (y/n)	if pre-existing version: Changes implemented by cApStAn?[2]	National domain experts in cApStAn translation process?	Change requests during verification (e.g. by NC or cApStAn)[2]	Changes imple-mented? [2]	Any comments after the Field Trial?[3]	Changes imple-mented? [3]	Fielded BFI-2 -XS version *identical* to?	Fielded BFI-2-XS version *similar* to?	
ara (Arabic)	ISR (Israel)	Only C, O	y	0	n	0 out of 12	0 out of 12	3	2			Changes for Main Study could also refer to items not fielded in the Field Trial.
cat (Catalan)	ESP (Spain)	BFI-2-S	y	10	n	0	0	0	0		val – ESP[4]	Catalan and Valencian are counted as the same language. BFI-2-S fielded in Main Study.
ces (Czech)	CZE (Czechia)	BFI-2-S	y	19	n	0	0	0	0			High number of changes aiming at gender-consistent wording. BFI-2-S fielded in Main Study. Changes in response options.
dan (Danish)	DNK (Denmark)	BFI-2-S	y	0	n	0	1	0	0			
deu (German)	DEU (Germany)	BFI-2-S	y	3	y	1	1	0	0	deu - AUT	deu – CHE[5]	BFI-2-S fielded in Main Study. Changes in response options.
deu (German)	AUT (Austria)	BFI-2-S	see DEU	see DEU	see DEU	1	1	0	0	deu - GER	deu – CHE[5]	Changes in response options

		Adaptations of the BFI-2-S for PIAAC Field Trial						Adaptations of the BFI-2-XS for PIAAC Main Study				
deu (German)	CHE (Switzerland)	BFI-2-S	see DEU	see DEU	see DEU	1	1	0	0		deu – DEU, AUT[5]	Changes in response options
eng (English)	SHP (Singapore)	BFI-2-S	y (master)	n	n	2	2	0	0	eng - NZL	eng – CAN, GBR, IRL[6]	
eng (English)	NZL (New Zealand)	BFI-2-S	y (master)	n	n	1	1	0	0	eng - SHP	eng – CAN, GBR, IRL[6]	BFI-2-S fielded in Main Study.
eng (English)	IRL (Ireland)	BFI-2-S	y (master)	n	n	0	0	0	0	eng - CAN, GBR	eng – NZL, SHP[6]	
eng (English)	CAN (Canada)	BFI-2-S	y (master)	n	n	1	1	0	0	eng - IRL, GBR	eng – NZL, SHP[6]	BFI-2-S fielded in Main Study.
eng (English)	GBR (Great Britain)	Only C, O	y (master)	n	n	0 out of 12	0 out of 12	0	0	eng - IRL, CAN	eng – NZL, SHP[6]	
esp (Spanish)	ESP (Spain)	BFI-2-S	y	5	n	0	0	0	0		esp – CHL[7]	BFI-2-S fielded in Main Study.
esp (Spanish)	CHL (Chile)	BFI-2-S	see ESP (esp)	9	n	3	15	0	0		esp – ESP[7]	High number of changes aiming at gender-consistent wording BFI-2-S fielded in Main Study.
est (Estonian)	EST (Estonia)	BFI-2-S	y	0	n	2	2	1	1			BFI-2-S fielded in Main Study. Changes in response options
fin (Finnish)	FIN (Finland)	BFI-2-S	n	n/a	y	1	1	4	4			Changes in response options
fra (French)	FRA (France)	BFI-2-S	y	0	n	4	4	5	5		fra – CHE, CAN[8]	

			Adaptations of the BFI-2-S for PIAAC Field Trial					Adaptations of the BFI-2-XS for PIAAC Main Study				
fra (French)	CHE (Switzerland)	BFI-2-S	see FRA	see FRA	see FRA	1	1	0	0		fra – FRA, CAN[8]	
fra (French)	CAN (Canada)	BFI-2-S	see FRA	see FRA	see FRA	7	7	1	1		fra – CHE, FRA[8]	BFI-2-S fielded in Main Study.
heb (Hebrew)	ISR (Israel)	Only C, O	y	1	n	0 out of 12	0 out of 12	4	4			Changes for Main Study could also refer to items not fielded in the Field Trial.
hrv (Croatian)	HRV (Croatia)	BFI-2-S	n	n/a	y	0	0	0	0			BFI-2-S fielded in Main Study.
hun (Hungarian)	SVK (Slovakia)	BFI-2-S	see HUN	see HUN	see HUN	0	0	0	0		hun – HUN[9]	BFI-2-S fielded in Main Study
hun (Hungarian)	HUN (Hungary)	Only C, O	n	n/a	y	3 out of 12	3 out of 12	0	0		hun – SVK[9]	
ita (Italian)	ITA (Italy)	BFI-2-S	n	n/a	n	13	13	0	0		ita – CHE[10]	High number of changes aimed at gender-consistent wording BFI-2-S fielded in Main Study.
ita (Italian)	CHE (Switzerland)	BFI-2-S	see ITA	see ITA	see ITA	0	0	0	0		ita – ITA[10]	
kor (Korean)	KOR (South Korea)	BFI-2-S	n	n/a	n	0	0	3	3			Korea did not field the BFI-2 XS nor the BFI-2 S in the Main Study. Changes in response options.
lav (Latvian)	LAV (Latvia)	BFI-2-S	n	n/a	n	1	1	0	0			
ltu (Lithuanian)	LIT (Lithuania)	Only C, O	n	n/a	n	1 out of 12	1 out of 12	1	1			Changes for Main Study could also refer to items not fielded in the Field Trial.

			Adaptations of the BFI-2-S for PIAAC Field Trial					Adaptations of the BFI-2-XS for PIAAC Main Study				
nld (Dutch)	NLD (Netherlands)	BFI-2-S	y	0	n	2	0	0	0		nld – BEL[11]	
nld (Dutch)	BEL (Belgium)	BFI-2-S	see NLD	see NLD	see NLD	1	1	0	0		nld – NLD[11]	
nob (Norwegian)	NOR (Norway)	BFI-2-S	y	0	n	5	5	1	1			BFI-2-S fielded in Main Study. Changes in response options
pol (Polish)	POL (Poland)	BFI-2-S	y	0	n	24	24	1	1			High number of changes, aimed mainly at gender-consistent wording and correcting typing errors Changes in response options
por (Portuguese)	PRT (Portugal)	BFI-2-S	n	n/a	n	0	14	0	0			High number of changes: Country changed translations without requesting changes. BFI-2-S fielded in Main Study.
rus (Russian)	EST (Estonia)	BFI-2-S	y	9	n	1	0	0	0	rus - LVA		BFI-2-S fielded in Main Study.
rus (Russian)	LVA (Latvia)	BFI-2-S	see EST (rus)	see EST (rus)	see EST (rus)	0	0	0	0	rus - EST		
slo (Slovakian)	SVK (Slovakia)	BFI-2-S	y	10	n	0	0	2	2			BFI-2-S fielded in Main Study. Changes in response options.
swe (Swedish)	SWE (Sweden)	BFI-2-S	n	n/a	y	0	0	1	1		swe – SWE[12]	Changes in response options

			Adaptations of the BFI-2-S for PIAAC Field Trial					Adaptations of the BFI-2-XS for PIAAC Main Study				
swe (Swedish)	FIN (Finland)	BFI-2-S	see SWE	see SWE	see SWE	0	11	2	8		swe – FIN[12]	High number of changes: Country changed translations without requesting changes. Changes in response options
val (Valencian)	ESP (Spain)	BFI-2-S	see ESP (Cat)	see ESP (Cat)	see ESP (Cat)	0	0	0	0		cat – ESP[4]	Catalan and Valencian are counted as the same language. BFI-2-S fielded in Main Study.

Note: If not otherwise indicated, the BFI-2-XS was fielded in Main Study. FT = Field Trial; NC = national centre; y = yes; n = no.

1 "Only C, O" indicates that only the six Conscientiousness and six Openness items from the BFI-2-S were administered in the Field Trial (i.e. 12 items in total).

2 Refers to the 30 items of the BFI-2-S, if not otherwise specified.

3 Refers to the 15 items of the BFI-2-XS. Thus, in cases where only selected domains were fielded, change requests could also refer to items not fielded in the Field Trial.

4 Similar versions: cat – ESP similar to val – ESP (9x language-specific suffixes; 1x different pronoun translation)

5 Similar versions: deu – DEU/AUT similar to deu – CHE (1x differently translated response options)

6 Similar versions: eng – CAN/GBR/IRL similar to eng – SHP/NZL (1x spelling difference)

7 Similar versions: esp – ESP similar to esp – CHL (3x different item formulations; 2x different adjectives; 3x different types of gendering)

8 Similar versions: fra – FRA similar to fra – CAN (5x different formulations; 3x different punctuation) and similar to fra – CHE (5x different formulations); fra – CHE similar to fra – CAN (1x different formulations; 3x different punctuation)

9 Similar versions: hun – HUN similar to hun – SVK (1x different adjective; 1x different item formulations)

10 Similar versions: ita – ITA similar to ita – CHE (5x different order gendering)

11 Similar versions: nld – NLD similar to nld – BEL (1x different item formulation)

12 Similar versions: swe – SWE similar to swe – FIN (2x different item formulation; 1x spelling difference)

6. Source instrument

The source instrument for the non-cognitive skills module in PIAAC 2023 was the Anglo-American original version of the BFI-2-XS (Soto and John, 2017[8]), which consists of 15 short-phrase items—three per Big Five domain. These items are answered on a 5-point rating scale ranging from *disagree strongly* (1) to *agree strongly* (5); the neutral category is labelled "neutral; no opinion" (3).

To facilitate translation and comparability, all 15 items of the BFI-2-XS were adapted slightly to form full sentences rather than phrases (see Online Appendix A on our OSF site; https://osf.io/u4f36/). The formulation of the response categories was also adapted slightly to fit the format typically applied in PIAAC. The adapted labels were "strongly disagree," "disagree," "neither agree nor disagree," "agree," and "strongly agree."

7. Translation procedure

State-of-the-art translation procedures were applied for all instruments assessed in PIAAC. The specific procedure applied for the translation of the non-cognitive skills module deviated from that applied for the other measures administered in PIAAC, in that the Big Five module was translated centrally by the PIAAC Consortium rather than by the individual participating countries. The intention was to take account of the sensitivity of the Big Five items to even slight meaning shifts and to ensure a monitored double translation and reconciliation design for these items. In preparation for the translations, the PIAAC Consortium (together with Christopher J. Soto, the first author of the BFI-2-XS) compiled a list of existing translations of the BFI-2-XS. These existing (and often already validated) versions were used as a source for the national adaptations of the BFI-2-XS to be used in PIAAC. Further, a list of one or more content experts for the Big Five in each of the 29 countries participating in the Big Five assessment was compiled, many of whom were authors of existing BFI-2 translations or at least similar personality inventories. These experts were later contacted to serve as expert reviewers for any new translations into the target languages.

8. Translation notes

To ensure item comparability across language versions and to prevent mistranslations, translation and adaptation notes (also known as item-specific guidelines or translation annotations) were provided for the BFI-2-S. These notes describe what specific words or phrases mean in measurement terms so that translators can transfer this meaning correctly without having to adhere too closely to the wording and structure of the source instrument. For the present purpose, the translation notes for the BFI-2 provided by Soto and John (personal communication, November 14, 2018) were used and expanded by cApStAn. For instance, the item "I am compassionate, have a soft heart" was accompanied by the note: "'Has a soft heart' means 'is caring and compassionate'."

The adapted translation notes for the BFI-2-S items are provided in Online Appendix B (https://osf.io/u4f36/).

9. Initial translated BFI-2-S versions

In a first step, it was checked whether a translation of the BFI-2-S into the target language already existed. This could be a version used in a previous OECD pilot study (Rammstedt, Lechner and Danner, 2024[9]) or a version from an independent translation project[3]. These translated versions were then used as the basis for the corresponding national adaptations. The respective translators, who were selected, trained, and supervised by cApStAn on behalf of the PIAAC Consortium, reviewed these preexisting translations and suggested edits if needed—for example, if the existing item translation did not conform to the translation notes. Revisions to the adaptations made to the BFI-2-S source version used for PIAAC were also carried out at this stage.

If no translation of the BFI-2-S into the target language previously existed, the PIAAC Consortium produced translations of the items following a double translation and reconciliation approach (Lyberg et al., 2021[10]), a slightly modified version of the TRAPD[4] procedure (Harkness, 2003[11]). Specifically, two independent translations into the target language were produced by professional translators with extensive experience in translating surveys and psychological assessments. These two versions were then reconciled into one translation by a senior questionnaire translator who merged them by (a) selecting the best components from each version; (b) selecting one version over the other; or, very rarely, (c) proposing a new version in case neither of the provided versions was deemed satisfactory. Problematic issues were discussed and resolved in a subsequent meeting between the initial translators and the reconciler. For all corresponding countries, scholars from the above-mentioned list of domain experts for the Big Five were then contacted. Whenever possible, the resulting translations were reviewed by these experts.

In the case of countries sharing an official language (e.g. German for Germany, Austria, and Switzerland), one language version (either preexisting or newly translated) served as the starting point for all countries and subsequently underwent further country-specific adaptation).

Regardless of the translation approach (use of an existing translation version or translation from scratch), countries were asked to review the translations provided by the PIAAC Consortium and request changes if problems were identified. These requests were reviewed by independent verifiers (linguists trained to identify potential equivalence issues in translated or adapted questionnaires) commissioned by cApStAn, and in some cases also by further members of the Consortium (e.g. domain experts). This step was called "verification." Based on this feedback, the initial translated versions were finalised and assessed in the PIAAC Field Trial.

All steps in the process from translation to finalisation of the instrument were rigorously documented, including the different translation versions at each step and additional comments by translators, verifiers, country teams, and domain experts if translation challenges or problems arose.

---

[3] For Germany, for example, see Rammstedt et al. (2020[16]).

[4] T: Translation, R: Review, A: Adjudication, P: Pretest, D: Documentation.
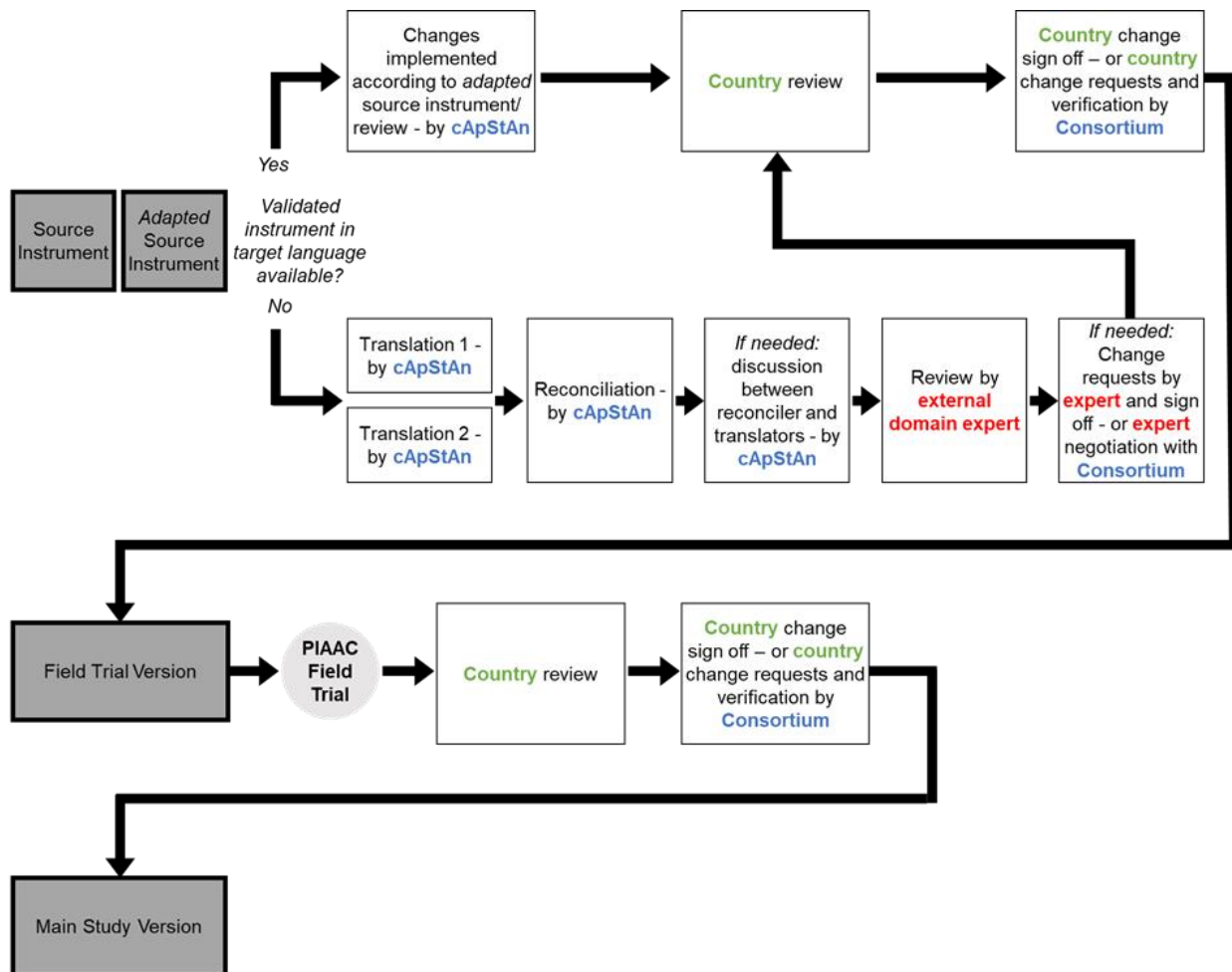
10. Final national versions of the BFI-2-XS

Data from the PIAAC Field Trial were analysed centrally by the PIAAC Consortium.[5] For the BFI-2-XS, the analyses focused on descriptive statistics and distributions of responses as well as dimensionality, scale reliability, and validity of the five domains. Besides overall inspection of predictive validity for several criteria, confirmatory factor analysis was used to inspect within each country the model fit for each dimension and for a joint 5-dimensional measurement model. In addition, multi-group confirmatory factor analysis was applied to inspect the cross-national functioning of the measurement of the five domains.

Countries received a common international report, as well as country-specific national reports. Based on an alignment procedure within multi-group confirmatory factor analysis, countries could identify any flagged severe national deviations from the international results. Based on these findings, countries were asked to review their respective item translations for potential translation biases.

Countries could then request revisions to their national BFI-2-S/BFI-2-XS versions from the PIAAC Consortium. The final BFI-2-XS/BFI-2-S versions administered in the PIAAC Main Study were thus either the initial versions used in the Field Trial or the slightly modified versions incorporating country feedback based on the Field Trial results. The general translation process is summarised in Figure 1.

---

[5] Data of the PIAAC Field Trial are confidential. Thus, no empirical findings can be reported here.

Figure 1. Translation and adaptation process for the BFI-2-S/BFI-2-XS followed in PIAAC

11. Results

Table 1 outlines for each of the 29 countries and 39 language versions the exact translation procedure and adaptation steps followed for the BFI-2-XS. The detailed procedure and adaptation steps per country version and item are described in Appendix C; the final 39 BFI-2-XS language versions are provided in Appendix A (for both see https://osf.io/u4f36/).

As described above, in PIAAC it was initially intended to assess only two of the Big Five dimensions – Conscientiousness and Openness – with six items each from the BFI-2-S. However, 24 countries decided to implement the full BFI-2-S in the PIAAC Field Trial. Thus, for these 24 countries translations and/or adaptations of the full BFI-2-S were prepared, resulting in 34 language versions. These are provided in Online Appendix A (see https://osf.io/u4f36/).

For 15 of the 24 languages in which the PIAAC non-cognitive skills module was administered, BFI-2-S/BFI-2-XS versions already existed, some of which had already been validated (see Table 2). These existing language versions were consulted and adapted for use in PIAAC. For all these country versions, the adaptations applied to the source version (see above) were implemented. In addition, the PIAAC Consortium sometimes also

changed item wording. These adaptations were often minor, for example, adding the feminine gender (as done in Czech) or correcting typing errors. However, in some cases, for example in the Spanish and Slovakian adaptations, items were reformulated to better represent the construct in the target language, align with translation notes, or capture common language use.

Table 2. Languages of the BFI-2-S/XS adaptations and sources of preexisting BFI-2(-S) versions

	Language	Pre-existing BFI-2(-S) version
1	Arabic	Translation was in progress; no publication has yet resulted.
2	Catalan/Valencian	Translation was in progress; no publication has yet resulted
3	Croatian	---
4	Czech	Hřebíčková et al. (2020[12])
5	Danish	Vedel et al. (2021[13])
6	Dutch	Denissen et al. (2019[14])
7	English	Soto and John (2017[8]; 2017[7])
8	Estonian	Translation was in progress; no publication has yet resulted.
9	Finnish	---
10	French	Translation from PIAAC Pilot was used.
11	German	Danner et al. (2019[15]); Rammstedt et al. (2020[16])
12	Hebrew	Translation was in progress; no publication has yet resulted.
13	Hungarian	---
14	Italian	---
15	Korean	---
16	Latvian	---
17	Lithuanian	---
18	Norwegian	Føllesdal and Soto (2022[17])
19	Polish	Translation from PIAAC Pilot was used; translation was in progress no publication has yet resulted.
20	Portuguese	---
21	Russian	Shchebetenko et al. (2019[18])
22	Slovakian	Halama et al. (2020[19])
23	Spanish	Gallardo-Pujol et al. (2022[20])
24	Swedish	---

For nine languages (or 12 language versions), there were no preexisting BFI-2-S adaptations. For these, the PIAAC Consortium conducted translations and, where possible, had them reviewed by Big Five experts in the respective countries.
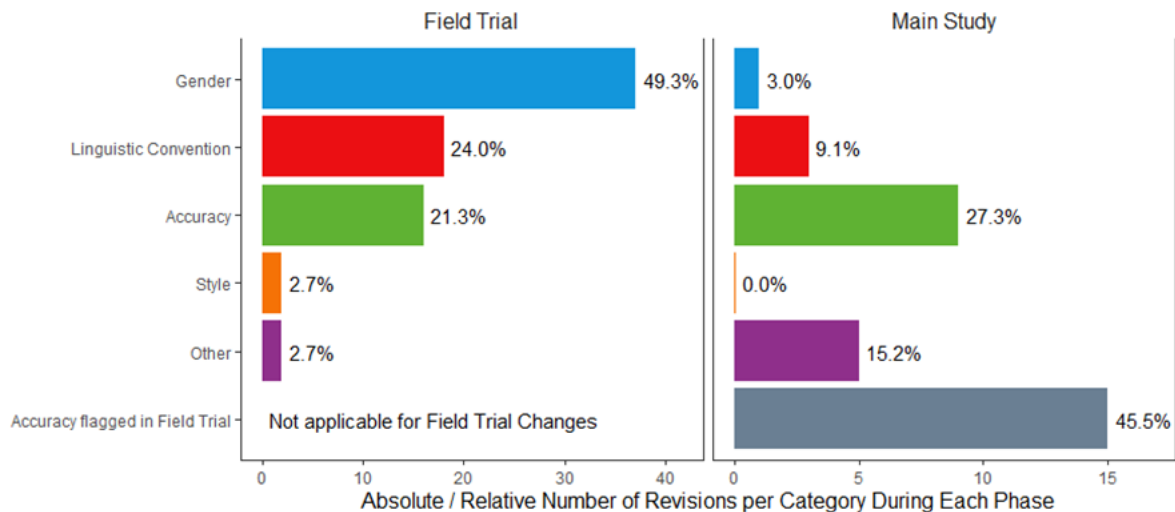
All language versions were thoroughly reviewed by the respective national centres in the verification phase prior to the Field Trial, and changes were requested when needed. These requests were then discussed within the PIAAC Consortium and implemented where appropriate.

For the purpose of this study, we coded the requests for changes—and thus the errors in the existing translations identified by the national centres—according to a customised MQM (Multidimensional Quality Metrics) error typology. MQM is a widely used framework for systematic translation quality assessment (https://themqm.org). We differentiated the following error types: (a) accuracy errors, (i.e. meaning-related errors); (b) accuracy errors as flagged in the Field Trial; (c) linguistic convention errors related to the linguistic correctness of the text (spelling, grammar, punctuation, etc.); (d) errors of style that reflect an inappropriate language use (awkward or unidiomatic text, etc.); (e) errors related to

survey-specific terminology that reflect inappropriate wording of response scales; (f) gender errors (i.e. missing or inconsistent gendering); and (g) other (e.g. changes to the English source wording or to the wording used in another country).

In the following, we will concentrate on requests for changes to item formulations. Some countries also requested that the labelling of the response scale options be modified. These cases are documented in Online Appendix C (see https://osf.io/u4f36/) only. As can be seen in the left panel of Figure 2 across all country versions, 49% ($n = 37$) of the requested and implemented changes to item formulations during the verification phase prior to the Field Trial related to adaptations of item gendering. Twenty-four percent ($n = 18$ requests) of the changes were due to linguistic conventions, and 21% ($n = 16$ requests) were due to accuracy of the item formulations. As an example of the latter, France requested that the preexisting French BFI-2 adaptation of the original item "I can be cold and uncaring" be changed. It was originally translated as "Je suis parfois dédaigneux/euse, méprisant(e)" [I am sometimes disdainful and contemptuous]. The national centre argued that "dédaigneux/euse" and "méprisant(e)" were stronger terms in French than "cold" and "uncaring" in the source item. This would have compromised the comparability of the translation with the other language versions, specifically regarding item difficulty. France therefore requested that the item be reformulated to "Je suis parfois indifférent(e), insensible" [I am sometimes indifferent, insensitive]. Averaged across all country versions, 10% of the BFI-2-S items were changed before the Field Trial, with changes per country version ranging between 0% and 80% of items.

Figure 2. Revisions to the translations of the BFI-2-XS before the Field Trial and the Main Study by Multidimensional Quality Metrics (MQM) category

The finalised BFI-2-S versions (or in some cases only two of the five trait domains) were then administered in a large-scale Field Trial in each country with a total sample size of nearly 30 000 respondents. The resulting data were analysed centrally by the PIAAC Consortium as described above. Based on these analyses, countries received detailed feedback on the psychometric performance of the national BFI-2-S adaptations. The psychometric evaluation comprised inter alia descriptive statistics on both the scale and the item level, scale reliability, exploratory and confirmatory factor analyses, and measurement invariance analyses across countries. Through these analyses, divergences from the

international results were flagged, and countries were asked to inspect these cases especially.

Based on this psychometric feedback, countries examined the item formulations, looked for specific causes of item misinterpretations, and suggested reformulations. The option to request changes was availed of for 13 of the 39 language versions. As can be seen in the right panel of Figure 2, nearly half of these changes were due to specific items being flagged in the Field Trial (46% or $n = 15$ requests).[6] For example, in the Finnish-language version for Finland, the item "I tend to be quiet" was flagged. The national centre reviewed the Finnish translation and concluded that the formulation used in the Field Trial – "Olen yleensä hiljainen" [I am usually quiet]—was stronger than the source formulation. They thus requested that the item wording be changed to "Olen usein hiljainen" [I am often quiet].

Other requests for changes were also due mostly to issues of accuracy in the item formulation—however, without the item issue being explicitly referred to in the documentation as having been flagged in the Field Trial (27% or $n = 9$ requests). Further, in a few cases, item translations were adapted to harmonise them with other versions of the same language or to correct typing errors. Such requests for changes were again reviewed by the PIAAC Consortium and almost always implemented. On average, 6% of the BFI-2-XS items per country version were changed at this point (with a range of 0% to 50%), resulting in the final BFI-2-XS versions administered in the PIAAC Main Study.

12. Discussion

In the present study, we presented adaptations of the BFI-2-XS developed as part of PIAAC Cycle 2 for 29 countries in 24 languages, resulting in a total of 39 language versions. New language versions were developed following a thorough and state-of-the-art procedure with two independent professional translations, reconciliations, expert consultations whenever possible, verifications, empirical pretesting, and, overall, an in-depth review and revision process. Preexisting translations similarly underwent linguistic reviews, verifications, empirical pretesting, review, and revision. All decisions taken in each individual step were documented to promote transparency and future use.

The translation and adaptation procedure was fully standardised and conducted in parallel with the aim of achieving fully comparable BFI-2-XS versions across all countries and languages. This approach of centrally producing adaptations in parallel based on a unified methodological framework and within a joint project structure has an advantage over independent individual translation projects because it has a higher likelihood of achieving comparable language versions. A lack of comparability could bias conclusions drawn from cross-cultural research.

The multi-step procedure described here, which included both domain experts and professional translators, was time and resource intensive. The results of the PIAAC Field Trial and Main Study indicate that this was time and money well spent, in that the procedure yielded brief measures of the Big Five traits that function well and are highly comparable across languages and cultures. However, the translation and adaptation process is always contingent on time constraints, and this is especially true in PIAAC. Unfortunately, these constraints did not allow all countries that share a language to harmonise their versions of

---

[6] The categorisation of the requested changes in the category "Accuracy flagged in Field Trial" is a conservative estimate, as we counted only those requests that explicitly referred to such flagging.

the instrument in an additional iteration. Such a step would have been beneficial in terms of comparability across these countries.

The aims of the present study were twofold. From a general point of view, we aimed to present this state-of-the-art translation and adaptation procedure: From a more pragmatic perspective, our aim was to share the different language versions of the BFI-2-XS developed in this process with the research community. We hope that other researchers can now use these translations in their own research, thereby promoting both within-country and international research across a broad range of languages and cultural contexts.

However, there is also a clear need for further research. Although all BFI-2-S and BFI-2-XS versions presented here were developed thoroughly, they have not yet been fully validated. The exact psychometric properties of each adaptation should thus be investigated in future studies and should be compared with those of the original English-language source version. Further, for languages in which alternative BFI-2 versions already existed, or in cases where the preexisting version was modified in the present adaptation process, future studies should compare these different adaptations and advise the research community about which version is best suited for research and applied use.

In sum, the present study describes the development and implementation of a coordinated procedure for developing, revising, and validating a psychological inventory in multiple languages and cultural contexts simultaneously. We hope that it will help enhance the quality of future translation projects, thereby increasing the comparability of the resulting language versions. Of equal importance, by sharing the resulting adaptations of the BFI-2-XS we also hope to support future cross-cultural research using the Big Five.

References

Danner, D. et al. (2019), "Das Big Five Inventar 2", *Diagnostica*, Vol. 65/3, pp. 121-132, https://doi.org/10.1026/0012-1924/a000218. [15]

Denissen, J. et al. (2019), "The Big Five Inventory–2: Replication of Psychometric Properties in a Dutch Adaptation and First Evidence for the Discriminant Predictive Validity of the Facet Scales", *Journal of Personality Assessment*, Vol. 102/3, pp. 309-324, https://doi.org/10.1080/00223891.2018.1539004. [14]

Føllesdal, H. and C. Soto (2022), "The Norwegian Adaptation of the Big Five Inventory-2", *Frontiers in Psychology*, Vol. 13, https://doi.org/10.3389/fpsyg.2022.858920. [17]

Gallardo-Pujol, D. et al. (2022), "Factor Structure, Gender Invariance, Measurement Properties, and Short Forms of the Spanish Adaptation of the Big Five Inventory-2", *Psychological Test Adaptation and Development*, Vol. 3/1, pp. 44-69, https://doi.org/10.1027/2698-1866/a000020. [20]

Halama, P. et al. (2020), "Slovak Adaptation of the Big Five Inventory (BFI-2): Psychometric Properties and Initial Validation", *Studia Psychologica*, Vol. 62/1, https://doi.org/10.31577/sp.2020.01.792. [19]

Harkness, J. (2003), "Questionnaire translation", in Harkness, J., F. van de Vijver and P. Mohler (eds.), *Cross-cultural survey methods*, Wiley. [11]

Hřebíčková, M. et al. (2020), "Pětifaktorový dotazník BFI-2: Hierarchický model s 15 subškálami. [Big Five Inventory 2 (BFI-2): Hierarchial model.]", *Československá Psychologie: Časopis Pro Psychologickou Teorii a Praxi*, Vol. 64/4, pp. 437-460. [12]

John, O., E. Donahue and R. Kentle (1991), "The Big Five Inventory: Versions 4a and 54", University of California, Institute of Personality and Social Research. [6]

John, O., L. Naumann and C. Soto (2008), "Paradigm Shift to the Big-Five Trait Taxonomy: History, measurement, and conceptual issues", in John, O., R. Robins and L. Pervin (eds.), *Handbook of personality: Theory and research*, Guilford, New York, NY. [1]

Levinsky, M., H. Litwin and C. Lechner (2019), "Personality traits: The ten-item Big Five Inventory (BFI-10)", in Bergmann, M., A. Scherpenzeel and A. Börsch-Supan (eds.), *SHARE Wave 7 Methodology: Panel innovations and life histories*, MEA, Max Planck Institute for Social Law and Psychology. [4]

Ludeke, S. and E. Larsen (2017), "Problems with the Big Five assessment in the World Values Survey", *Personality and Individual Differences*, Vol. 112, pp. 103-105, https://doi.org/10.1016/j.paid.2017.02.042. [3]

Lyberg, L. et al. (2021), *AAPOR/WAPOR task force report on quality in comparative surveys*. [10]

McCrae, R. and P. Costa (2008), "Empirical and Theoretical Status of the Five-Factor Model of Personality Traits", in Boyle, G., G. Matthews and D. Saklofske (eds.), *The SAGE Handbook of Personality Theory and Assessment: Volume 1 — Personality Theories and Models*, SAGE Publications Ltd, London, https://doi.org/10.4135/9781849200462.n13. [2]

Rammstedt, B. et al. (2020), "Validation of the Short and Extra-Short Forms of the Big Five Inventory-2 (BFI-2) and Their German Adaptations", *European Journal of Psychological Assessment*, Vol. 36/1, pp. 149-161, https://doi.org/10.1027/1015-5759/a000481. [16]

Rammstedt, B. and O. John (2007), "Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German", *Journal of Research in Personality*, Vol. 41/1, pp. 203-212, https://doi.org/10.1016/j.jrp.2006.02.001. [5]

Rammstedt, B., C. Lechner and D. Danner (2024), "Beyond Literacy: The Incremental Value of Non-Cognitive Skills"*, OECD Education Working Papers*, OECD Publishing, Paris. [9]

Rammstedt, B., L. Roemer and C. Lechner (2023), "Consistency of the Structural Properties of the BFI-10 Across 16 Samples From Eight Large-Scale Surveys in Germany", *European Journal of Psychological Assessment*, https://doi.org/10.1027/1015-5759/a000765. [21]

Shchebetenko, S. et al. (2019), "Measurement Invariance and Sex and Age Differences of the Big Five Inventory–2: Evidence From the Russian Version", *Assessment*, Vol. 27/3, pp. 472-486, https://doi.org/10.1177/1073191119860901. [18]

Soto, C. and O. John (2017), "Short and extra-short forms of the Big Five Inventory–2: The BFI-2-S and BFI-2-XS", *Journal of Research in Personality*, Vol. 68, pp. 69-81, https://doi.org/10.1016/j.jrp.2017.02.004. [8]

Soto, C. and O. John (2017), "The next Big Five Inventory (BFI-2): Developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power.", *Journal of Personality and Social Psychology*, Vol. 113/1, pp. 117-143, https://doi.org/10.1037/pspp0000096. [7]

Vedel, A. et al. (2021), "Development and Validation of the Danish Big Five Inventory-2", *European Journal of Psychological Assessment*, Vol. 37/1, pp. 42-51, https://doi.org/10.1027/1015-5759/a000570. [13]