MEASUREMENT PROPERTIES OF TRANSLATED VERSIONS OF MEASURES ASSESSING THE SUBJECTIVE EFFECTS OF TOBACCO- AND NICOTINE-CONTAINING PRODUCTS: A LITERATURE REVIEW

Acquadro C[1], Desvigne-Gleizes C[2], Mainy N[3], Arnould B[1], Spies E[3], Hankins M[3], Weitkunat R[3], Chrea C[3]

[1]Mapi, An ICON plc Company, Lyon, France; [2]Mapi Research Trust, Lyon, France; [3]PMI R&D, Philip Morris Products S.A., Neuchâtel, Switzerland

Background and Objectives

- Several measures have been widely used to assess dependence, craving, withdrawal symptoms, and reinforcing effects in users of tobacco- and nicotine-containing products (TNP). These include: the Fagerström Test for Nicotine Dependence (FTND), the Questionnaire on Smoking Urges, original (QSU) and brief (QSU-b) versions, the Minnesota Nicotine Withdrawal Scale, original (MNWS) and revised (MNWS-R) versions, and the Cigarette Evaluation Questionnaire (CEQ) and modified version (mCEQ). While they have been translated into various languages, the translations and corresponding measurement properties have not been reviewed systematically.
- This study aimed at:
  1) Identifying the translations of the FTND, the QSU/QSU-b, the MNSW/MNWS-R, and CEQ/mCEQ for which psychometric properties are available, and the methods used to develop the translations
  2) Describing those properties and the context in which they were evaluated (e.g., target population and TNPs used by the population under study).

Methods

Search Strategy

We searched Medline and Embase (March 2018), with no limitation in timeframe, using the following key words:
1) Translation, Language, Version, Cross-cultural valid*, internal consistency, Cronbach alpha, reliability, Validation, responsiveness, validity; combined with: 2) QSU, Questionnaire on smoking urges, Cigarette Evaluation Questionnaire, Fagerström Test for Nicotine Dependence, FTND, Fagerström Test for Cigarette Dependence, Minnesota Nicotine Withdrawing Scale, MNWS. The combination of #1 AND #2 was limited to Abstract, Human research, and English. Reference lists were screened to identify additional relevant studies.

Selection criteria

Abstracts retrieved through the search strategy were reviewed and excluded, if they: 1) did not refer to the instruments of interest, 2) referred to the original U.S. version of the instruments of interest, or 3) did not include psychometric data. Conference abstracts were not retained. When a review paper was retrieved, references of interest cited in the article were added to the list of articles to be considered. Articles were included if they reported at least one of the measurement properties listed in the next section.

Methods (cont')

Measurement Properties

Using, as a basis, the **CO**nsensus-based **S**tandards for the selection of health **M**easurement **IN**struments (COSMIN) taxonomy and definitions, the measurement properties were categorized as follows: reliability, validity, and responsiveness[1].

Reliability is defined as the extent to which scores for respondents who have not changed are the same for measurement under several conditions (e.g., using different sets of items from the same instrument (internal consistency), over time (test-retest), by different persons on the same occasion (inter-rater), or by the same persons on different occasions (intra-rater)).

Validity is the extent to which an instrument measures the construct it is supposed to measure and contains the following measurement properties:

- Content validity: The degree to which the content of an instrument is an adequate reflection of the construct to be measured. As this review deals with translations, this part will include a description of the translation process and whether or not, on the qualitative level, the content of some items was changed to reflect cultural aspects.
- Construct validity: The degree to which the scores of an instrument are consistent with hypotheses *(i.e., with regard to internal relationships, relationships to scores of other instruments, or differences between relevant groups)* based on the assumption that the instrument validly measures the construct to be measured. Besides structural validity and hypothesis testing, cross-cultural validity has been included. Cross-cultural validity is the degree to which the performance of the items on a translated or culturally adapted instrument are an adequate reflection of the performance of the items of the original version of the instrument. This is assessed by means of multi-group factor analysis or differential item functioning (DIF) using data from a population that completed the questionnaire in the original language, as well as data from a population that completed the questionnaire in the new language.
- Predictive validity (is the considered instrument score predictive of a future outcome/event?), sensitivity, and specificity. The latter were included, as some of the instruments investigated were explored as screening tools.

Responsiveness is the ability of an instrument to detect change over time in the construct to be measured. Responsiveness is considered an aspect of validity in a longitudinal context.

Reference

1. Mokkink LB, et al. J Clin Epidemiol. 2010 Jul;63(7):737-45. https://www.cosmin.nl/tools/cosmin-taxonomy-measurement-properties/

Results

- The search retrieved 193 references. We selected 47 articles for data extraction, as they described measurement properties of translated versions of the FTND, QSU/QSU-b, and MNWS/MNWS-R [50-97]. No references were found about the CEQ or the mCEQ.
  See Tables 1 to 3 for the list of studies and translations.

Table 1. List of the FTND translations and corresponding studies

Language	Country	References
Arabic	Lebanon	1. Salameh et al. Addictive Behaviors. 2013;38(5):2174-9.; 2. Salameh et al. International journal of behavioral medicine. 2014;21(2):385-93.
Arabic	UK (Yemenite immigrants)	3. Kassim et al. Asian Pacific Journal of Cancer Prevention: Apjcp. 2012;13(4):1285-8. [3]
Arabic	Yemen	4. Nakajima et al. Journal of Psychoactive Drugs. 2012;44(5):437-41.
Chinese	Taiwan	5. Huang et al. Addictive Behaviors. 2006;31(12):2324-7; 6. Huang et al. J Clin Nurs. 2008 Apr;17(7):884-90.
Chinese	USA	7. Yamada et al. Addictive Behaviors. 2009;34(2):125-33.
Dutch	The Netherlands	8. Breteler et al. Addict Behav. 2004;29(1):199-205.
Dutch	The Netherlands	9. Vink et al. Addict Behav. 2005;30(3):575-9.
Farsi	Iran	10. Sarbandi et al. Journal of Research and Health 2015; 5(1): 96-103.
Farsi	Iran	11. Robabeh et al. NeuroQuantology. 2017;15(2):253-60.
French	France	12. Chabrol et al. Addictive Behaviors. 2003;28(8):1441-5; 13. Chabrol et al. Addict Behav. 2005 Aug;30(7):1474-7.
French	Switzerland	14. Etter et al. Addiction. 1999 Feb;94(2):269-81; 15. Etter et al. Drug and Alcohol Dependence. 2005;77(3):259-68.
German	Germany	16. John et al. Addict Behav. 2004;29(6):1207-12.
Hindi	India	17. Janghee et al. Nicotine and Tobacco Research. 2010;12(11):1162-6.
Italian	Italy	18. Ferketich et al. Psychological Reports. 2008;102(3):687-94; 19. Grassi et al. Psychological Reports. 2014;114(1):1-13; 20. Svicher et al. Addictive Behaviors. 2018;77:38-46.
Japanese	Japan	21. Mikami et al. Jpn J Cancer Res. 1999;90(10):1071-5; 22. Kawada et al. Work (Reading, Mass). 2010;35(2):183-9.
Korean	Korea	23. Park et al. Drug and Alcohol Dependence. 2004;74(2):197-203.
Malay	Malaysia	24. Yee et al. Malaysian Journal of Psychiatry 2011;20(1).
Mayalayam	India	25. Jayakrishnan et al. Asian Pacific Journal of Cancer Prevention: Apjcp. 2012;13(6):2663-7.
Norway	Norwegian	26. Stavem et al. Addiction. 2008;103(9):1441-9.
Portuguese	Brazil	27. Carmo et al. Rev Bras Med. 2002;59(1/2):73-80; 28. De Lima Osorio et al. Perspectives in Psychiatric Care. 2013;49(1):5-12.
Portuguese	Brazil	29. De Meneses Gaya et al. Nicotine Tob Res. 2009 Oct;11(10):1160-5.
Spanish	Spain	30. Becoña et al. Psychological Reports 1998;83:1455-8; 31. Becoña et al. Spanish Journal of Psychology. 2010;13(2):951-60
Spanish	Mexico	32. Moreno-Coutiño et al. Journal of Addictions Nursing. 2017;28(1):27-33.
Thai	Thailand	33. Klinsophon et al. Journal of the Medical Association of Thailand. 2017;100(10):1130-4.
Turkish	Turkey	34. Uysal et al. Tuberkuloz ve Toraks. 2004;52(2):115-21; 35. Uysal et al. Tuberkuloz ve Toraks. 2015;63(4):250-6.

Table 2. List of the QSU/QSU-b translations and corresponding studies

Measure	Language	Country	References
QSU-32	French	France	36. Guillin et al. Encephale. 2000;26(6):27-31.
	German	Germany	37. Muller et al. Zeitschrift fur Klinische Psychologie und Psychotherapie. 2001;30(3):164-71.
	Portuguese	Brazil	38. Araujo et al. Cad Saude Publica. 2006 Oct;22(10):2157-67.
	Spain	Spanish	39. Cepeda-Benito et al. Assessment 2004;11:152–9.
QSU-12	French	Belgium	40. Dethier et al. Neuropsychiatr Dis Treat. 2014;10:1459–68.
QSU-10 (QSU-b)	Chinese	China	41. Yu et al. Nicotine & Tobacco Research. 2010;12(6):630-4.
	Dutch	The Netherlands	42. Littel et al. Netherlands Journal of Psychology 2011;66:44-9.
	Malay	Malaysia	43. Blebil et al. J Epidemiol Glob Health. 2015;5(1):15-22.
	Spanish	Spain	44. Cepeda-Benito et al., Psychol Assess. 2004;16(4):402-7.

Table 3. List of the MNWS/MNWS-R translations and corresponding studies

Measure	Language	Country	References
MNWS 9-item version	Chinese	China	41. Yu et al. Nicotine & Tobacco Research. 2010;12(6):630-4.
	Korean	USA	45. Kim et al. Journal of Nursing Measurement. 2007;15(2):121-32.
	Malay	Malaysia	46. Blebil et al. Value in Health Regional Issues. 2014;3(1):19-23.
MNWS-R and MNWS 8-item version	Italian	Italy	47. Svicher et al. European Addiction Research. 2017;23(3):157-62.

- 60% of the translations (15 out of 25) were documented with a description of the translation process used to develop each of them. Out of these 15 translations, only three [3, 24, 34] were presented with a brief report of the difficulties encountered and solutions found. References to guidelines or recommendations were given for six translations [3, 7, 10, 24, 30, 33]. Descriptions of the translation process were either minimal with only a mention of the steps performed [4, 11, 17, 21, 25, 30] or more detailed with information about the people involved in the process [1-3, 7, 10, 14, 33, 24, 33, 34]. Except for one translation (i.e., French for Switzerland [14]), all teams included in their process a backward translation step (i.e., the translation of the target language version back to the source language, English).
- Classical Test Theory (CTT) was used for all translations, except for the Chinese version developed for immigrants to the U.S. [7], where only an item response theory (IRT) based approach was used. IRT supplemented by CTT was applied for the Dutch [8], and Italian version [20].
- Measurement equivalence was explored for only one translation (Chinese for immigrants to the U.S. [7]), where DIF was examined using IRT. Q2 (difficult to refrain) showed a significant DIF, suggesting respondents using the Chinese version were more likely to endorse this item and report more difficulty to refrain from smoking at various public places even after controlling for the nicotine dependence level. Interestingly, this DIF item in the Chinese version contributed minimally at the aggregate level. The impact of the DIF was negligible on scale scores. Both unidimensional and multidimensional results did not show DIF for Q1 (time to first cigarette) and Q3 (cigarette hated most to give up), suggesting that these two items are DIF-free. This DIF-free finding suggests that these two items could be kept in the FTND to allow a fair comparison between English- and Chinese-speaking smokers.
- Internal consistency was explored for all translations with Cronbach's alpha ranging from 0.52 (Thai version [33]) to 0.83 (Portuguese for Brazil [29]). Structural validity was documented for 72% of the translations (18 out of 25). 40% of the translations (10 out of 25) have been assessed for test-retest reliability. Concurrent validity was explored for 32% of the translations (8 out of 25). Predictive validity was investigated for 12% of the translations (3 out of 25). Sensitivity/specificity was explored for 24% of the translations (6 out of 25).
- Most of the FTND translations have properties similar to the original (data not shown).
- Table 4 presents the results of a selection of FTND studies.

We will focus our presentation on the FTND results

- The search retrieved 35 FTND studies. We identified 25 different FTND translations (see Table 1). In case of different studies exploring the same language (e.g., Japanese reported in two independent studies, or French reported in French and Swiss studies), we contacted the authors for clarification about the version used. Some authors confirmed the use of an existing translation [22] and others the use of their own version [9, 12, 13, 29, 32]. When we did not receive any answer, we counted only one version.
- Conventional cigarettes were the TNP evaluated in all studies. Bidis (locally made cigarettes by wrapping coarse tobacco in dried temburni leaf) were used in India [17, 25].
- A wide range of populations was investigated, of various levels of age and cigarette consumption (from light smokers to heavy smokers).

Table 4. Sample characteristics and psychometric properties of a selection of FTND translations and studies

Language / Country Study	Sample characteristics			Translation process described in paper Y/N	Reliability		Validity			
	N Gender M/F (%)	Mean age in years ±SD	Mean cig. Consumption/day ±SD		Internal consistency	Reliability (test-retest, inter-rater, intra-rater)	Structural validity	Hypothesis Testing Known Group validity	Hypothesis Testing Concurrent validity	Predictive validity Sensitivity, Specificity
Chinese / Taiwan Huang et al., 2006 [5]	245 97/3	47.0 ±16.6	20.0 ±10.6	**N**	α=0.74		Two factors F1:Q3, Q5 F2: Q1, Q2, Q4, Q6	Sa. Cotinine: r=0.45 (p<0.001) with Q5 r=0.07 ns		
Huang et al., 2008 [6]	250 95/5	47.6 ±17.1	47.7 ±17.2	**N**						Sa. Cotinine: cut-off score= +4; Se: 76.2%; Sp: 67.5% CO: cut-off score = +4; Se: 74.9%; Sp: 71.6%
French / France Chabrol et al., 2003 [12]	772 54/46	M: 37.2 ±9.3 F: 35.1 ±8.4	NS (only FTND scores Q4)	**N**	α=0.86 with Q3 excluded		One factor with exclusion of Q3			
Italian / Italy Ferketich et al., 2008 [18]	593 60/40	48.6 ±11.6	22.7 ±9.0	**N**	α=0.55 (M, α=0.59; F, α=0.50)			CO: r=0.27 (p<0.001) [M, r=0.27, F, r=0.25; p<0.001]		Prediction of abstinence: significant (p=0.009) at 7 wk, (OR 0.39, 95% CI 0.22-0.97); ns at 12 mo. (OR 0.92, 95% CI 0.83-1.01).
Korean / Korea Park et al., 2004 [23]	268 93/7	46.6 ±9.2	18.8 ±9.3	**Y**	α=0.72		Two factors F1: Q1, Q2, Q4, Q6 F2: Q3, Q5	Ur. Cotinine: r=0.49 (p<0.05) Duration of smoking: no correlation (p=0.08) Age at starting smoking: p<0.05, OR = 11.25, 95% CI 1.46–97.10 for initiation at 19 years or earlier/ 25 years or later		
Mayalayam / India Jayakrishnan et al., 2012 [25]	474 (intervention) 100/0 454 (control) 100/0	44.6 ±9.7 44.5 ±10.3	13.2 ±8.4 10.9 ±6.8	**Y**	α=0.70 (on a sample of n=170)	Test-retest . ICC=0.77 (95% CI: 0.67-84, (p<0.001) . 2 months (on a sample of n=91)		# packs smoked: r= 0.677 (p<0.001) Age reg. smoking: r=0.187 (p=0.022)		
Norway / Norwegian Stavem et al., 2008 [26]	292 (54/46)	42.0 ±15.0	12.5 ±5.9	**N**	α=0.61	Test-retest . ICC=0.90 (95% CI: 0.78-0.96) . 15 days (5-28 days) (on a convenience sample of n=31)		Age of reg. smoking: r=-0.22 (p<0.01) # pack-years smoked: r=0.44 (p<0.01) Willingness to pay for 1 cig. after 1 day without smoking: r=0.36 (p<0.01)	CDS-12: r=0.60 (p<0.001) CDS-5: r=0.72 (p<0.001)	
Turkish / Turkey Uysal et al., 2004 [34]	169 62/38	38.0 ±12.0	22.0 ±8.2	**Y**	α=0.56 (0.65 if Q3 omitted)	Test-retest . r=0.55 to 0.90 (p<0.01) – r provided for individual Q. Not provided for sum score	Two factors F1: Q1, Q2, Q4, Q5, Q6 F2: Q3			

α: Cronbach's alpha; CDS: Cigarette Dependence Scale; Cig.: Cigarettes; CP: Cigarette packs; CO: Carbon monoxide (exhaled); F: Female; F1: Factor 1; F2: Factor 2; ICC: Intra-class correlation coefficient; M: Male; mo: months; ns: Not significant; NS: Not specified; OR: Odds ratio; Q: Question; Reg.: Regular; Sa: Salivary; Se: Sensitivity; Sp: Specificity; Ur.: Urinary; wk: weeks; Y/N: Yes/No

Discussion

Translations: Number Retrieved and Translation Process

- **Given the globalization of tobacco research and control, we expected to retrieve more** than 25, nine, three, and one **translations** of the FTND, QSU/QSU-b, MNWS and MNWS-R, respectively, **documented with measurement properties.** A search on PROQOLID (https://eprovide.mapi-trust.org/) reveals, for instance, that there are 19 translations available for the QSU-b. Among those, we found two versions overlapping with our research (i.e., Dutch and Spanish (Spain)), indicating that there are two versions of the QSU-b in those languages. PROQOLID does not mention whether those 19 translations have undergone any evaluation of their measurement properties.
- **Our review showed that the translation process used is not standardized and not always documented.** This could prove to be a challenge if the Center for Tobacco Products aligns any future guidance with the 2009 patient-reported outcome (PRO) guidance published by the U.S. Food and Drug Administration (FDA) Center for Drug Evaluation and Research. The 2009 PRO guidance (Appendix VIII) outlines that all translation documents should be provided for FDA review. This includes a report on the process(es) used and challenges encountered during the translation process.

Psychometric Properties

- The TNPs used during the translation validation studies are **always conventional cigarettes.** None of the translations have been validated with alternative products, in particular smoke-free products, which are at the center of the public health debate on tobacco harm reduction.
- **There is a great heterogeneity** in the populations recruited for each study, in terms of sample characteristics (i.e., gender (samples with mixed genders or a majority of male subjects), age or level of cigarette consumption (light to heavy smokers)). In addition, not all properties are explored for each language depending on the objectives of the research teams.
- **Cross-cultural validity is rarely explored.** Measurement equivalence using an IRT based approach examining DIF is almost never applied. This raises a concern about the comparability of these measures across languages and cultures.

These results are showing:

1) Discrepancies between the number of translations available, with and without documented information about their measurement properties,
2) Heterogeneity in the scope of measurement properties explored, and in the characteristics of the samples recruited, and
3) Lack of validation with TNPs other than conventional cigarettes,

and raise the need for:

- **Implementing a centralized repository for measurement instruments.** Such a structure would enable researchers to have access to the most up-to-date information about measures, would prevent the development of multiple translations for the same language, and would enhance the integrity of measurement instruments.
- **Conducting validation studies** with alternative products, in particular smoke-free products.

PMI SCIENCE
PHILIP MORRIS INTERNATIONAL

ISPOR Europe 2018, Barcelona

10-14 November 2018

Competing Financial Interest

The research described in this poster was sponsored by Philip Morris International