



Validation of the Italian version of the In-Person Telephone Interview for Cognitive Status (IP-TICS)

Edoardo Nicolò Aiello¹ · Elena Guidetti² · Giulia De Luca¹ · Francesca Frisco¹ · Beatrice Curti¹ · Arianna Moreschi¹ · Lorenzo Esposti² · Sara Casartelli² · Chiara Curatoli¹ · Alice Zanin¹ · Elisa Camporeale¹ · Martina Andrea Sirtori^{2,3} · Federico Verde¹ · Vincenzo Silani^{1,4} · Nicola Ticozzi^{1,4} · Barbara Poletti^{1,5} · Nadia Bolognini^{2,6}

Received: 30 April 2026 / Accepted: 21 June 2026
© The Author(s) 2026

Abstract

Background This study validated the Italian version of the In-Person Telephone Interview for Cognitive Status (IP-TICS). **Methods** 468 healthy Italian individuals (176 males; age: 57.5 ± 15.4 , education: 13.2 ± 3.7) completed the TICS, IP-TICS, MMSE and MoCA. Correlational analyses explored the construct validity of the IP-TICS against the TICS, MMSE and MoCA. Hierarchical Confirmatory Factor Analyses tested the configural, metric and scalar invariance between the TICS and IP-TICS. Statistical equivalence was assessed *via* a two one-sided test (TOST) procedure. Equating norms were derived *via* log-linear smoothing equipercentile equating (LSEE) analyses, with TOST procedures assessing equivalence between empirical and LSEE-derived scores. **Results** The IP-TICS converged with the TICS, MMSE and MoCA, showing configural and metric invariance and equivalence with the TICS. Empirical and LSEE-derived scores were statistically equivalent. **Discussion** Findings support the convergence, equivalence and invariance between the TICS and the IP-TICS. The IP-TICS converges with established in-person screeners.

Keywords Telephone Interview for Cognitive Status · Cognitive screening · Tele-neuropsychology · Mini-Mental State Examination · Montreal Cognitive Assessment

Introduction

Telephone-based cognitive screening (TBCS) tests represent a flexible and acceptable alternative to deliver first-level cognitive assessments whenever topographical, logistic, and socio-economic barriers prevent examinees from accessing in-person facilities, therefore also allowing for a greater continuity of care in longitudinal scenarios [1, 2]. These instruments also serve research purposes, as facilitating the implementation and completion of large-scale, decentralized both observational and interventional studies [3]. When developing TBCS tools, it is crucial to establish whether they assess cognitive constructs comparable to traditional face-to-face measures, thereby ensuring that both administration modalities can be validly interchangeable [4, 5]. One approach to address this issue is test equating, which generates cross-walks between telephone- and

in-person-based measures to support score comparability across formats [6, 7], being particularly useful when the two measures at hand are inherently different (e.g., in terms of items and range) [6]. Alternatively, evidence can be provided that the same instrument can be administered across modalities while minimizing format-related adaptations [8]. This is, for instance, the case of the recently standardized Global Examination of Mental State [9], a novel cognitive screener for which a telephone-based counterpart has been developed [10] that features the same range and includes similar items.

Unsurprisingly, efforts devoted to these goals also targeted the Telephone Interview for Cognitive Status (TICS) [11] – which arguably represents the most widespread TBCS test for global cognition, whose clinimetric soundness and optimal usability have been thoroughly established [1–3]. In fact, several algorithms have been developed worldwide

Extended author information available on the last page of the article

that convert TICS scores to those on common in-person screening tests – such as the Mini-Mental State Examination (MMSE) [12] –, and vice-versa [6].

In this respect, and with specific regard to the Italian scenario, a key contribution lies in the 25-plus-year-old report by Ferrucci et al. [13], who explored whether the TICS, if administered face-to-face, could represent a feasible alternative to the MMSE. Within this report, the “In-Person TICS” (IP-TICS), which retained the content and scoring structure of the original TICS, showed excellent test-retest reliability, strong convergence with the MMSE, and comparable diagnostic accuracy for dementia [13]. Although Ferrucci et al. [13] originally proposed the IP-TICS as an alternative to the MMSE for individuals with sensory or motor limitations, their work also provided a proof of concept for the use of the TICS in face-to-face settings. Thus, a further, relevant follow-up to Ferrucci et al.’s [13] study would be represented by a full validation of the IP-TICS, as their report preceded the availability of the official Italian version of the TICS in 2021 [14, 15], addressed a relatively small ($N = 104$) and heterogeneous sample – including both patients with dementia (~ 40%) and healthy individuals – and did not directly compare the measurement properties of the IP-TICS with those of its original format. For such an investigation to be run properly, some key points should be considered. First, the convergence between the TICS and the IP-TICS should be established, and the measurement invariance across the two modalities demonstrated. In this regard, a further *desideratum* would be to determine the statistical equivalence between the TICS and the IP-TICS. Finally, it should be ascertained that the IP-TICS taps into the same cognitive construct captured by traditional in-person screening tests for global cognition, such as the MMSE and the Montreal Cognitive Assessment (MoCA) [16] – as recently demonstrated in respect to the TICS [17].

Beyond methodological considerations, validating an in-person adaptation of the TICS addresses a practical clinical need by determining whether the same cognitive screener can be reliably administered both over the telephone and *de visu*. Additionally, although TICS procedures provide a practical alternative when face-to-face assessments are not feasible, routine clinical practice often includes scenarios in which traditional in-person screeners may not fully address assessment needs. As suggested by Ferrucci et al. [13], a validated version of the IP-TICS may therefore represent a useful alternative when standard screeners such as the MMSE and MoCA are difficult to administer because of visual or motor limitations, bedside or isolation settings, or other contextual constraints affecting visuospatial or manual tasks. In these circumstances, preserving the reduced visuospatial and motor demands of the original TICS while maintaining compatibility with telephone-based administrations

may facilitate continuity and flexibility in cognitive screening, including longitudinal pathways alternating telephone and in-person assessments.

Given the above premises, this study aimed at validating an Italian version of the IP-TICS by testing its convergence, measurement invariance and equivalence with the TICS, and exploring its association with widespread in-person cognitive screening test – i.e., the MMSE and the MoCA. To further support the usability of the IP-TICS, an additional scope of this report was to derive equating norms to convert the TICS to the IP-TICS, and vice-versa.

Methods

Participants

A total of 481 healthy Italian individuals were initially recruited between 2022 and 2024 in accordance with a non-probability sampling procedure, namely *via* Authors’ personal acquaintances and cognitive screening campaigns carried out at our institution (details blinded for peer review). Exclusion criteria included: (1) formally diagnosed neurological or psychiatric disorders; (2) history of neurodevelopmental disorders; (3) severe or uncompensated medical conditions; (4) ongoing treatment with psychotropic medications; (5) severe and/or uncorrected hearing and/or visual defects. Additionally, participants were screened *via* the MMSE [18], and those obtaining a below-cutoff age- and education-adjusted score were further excluded to rule out sub-clinical, undetected cognitive deficits.

As a result, the final sample included 468 participants, whose demographic stratification is reported in Supplementary Table 1.

The study was approved by the institutional ethics committee (details blinded for peer review) and all participants provided written informed consent.

Materials

All participants were administered the Italian versions of the TICS [14, 15], the MMSE [18], the MoCA [19] the IP-TICS [13–15]. The Italian version of the TICS ranges from 1 to 41 and assesses *Orientation* (TICS-O, *range* = 0–12), *Language* (TICS-L, *range* = 1–8), *Memory* (TICS-M, *range* = 0–12) and *Attention and Executive Functioning* (TICS-AEF, *range* = 0–9). The additional *Delayed Recall* task (TICS-DR; *range* = 0–10) was also administered to compute a further total score, the “TICS_{&DR}” (*range* = 1–51). The IP-TICS mirrors the TICS in terms of item groupings and total/subscale-level ranges, with the following, minor procedural adjustments being made: (1) the task assessing

comprehension by tapping on the phone was modified so that the examinee was instructed to tap on the table instead – as in Ferrucci et al. [13].; (2) spatial orientation items requiring to provide highly specific information (e.g., street number) were substituted with more general questions (e.g., the Country where the examinee was in that moment). We deemed this last adaptation as necessary since the IP-TICS was intended to be administered in-person, and not at the examinee's place – as is typically the case for the TICS. In fact, requesting the examinee to know the details of the address of a place that is likely new to him/her (e.g., the clinic) would have led to an underestimation of his/her spatial orientation skills.

MoCA items were grouped in accordance with Aiello et al.'s [19] subdivision, and more specifically: MoCA - *Visuo-Spatial* (MoCA-VS; range = 0–4), *Executive Functioning* (MoCA-EF; range = 0–4), *Language* (MoCA-L; range = 0–5), *Attention* (MoCA-A; range = 0–6), *Memory* (MoCA-M; range = 0–5) and *Orientation* (MoCA-O; range = 0–6). As to the MMSE, the following sub-scores were computed (as detailed elsewhere [17]): *Orientation* (MMSE-O; range = 0–10); *Attention* (MMSE-A; range = 0–5); *Memory* (MMSE-M; range = 0–6); *Language* (MMSE-L; range = 0–8). The *Constructional praxis* item was left out of these groupings and thus accounted for only within its total score.

Procedure

Data collection was conducted by licensed psychologists and psychology trainees that underwent an extensive training focused on standardized test administration procedures. Participants underwent both in-person and telephone-based assessments. MMSE, MoCA, and IP-TICS were administered in person, while TICS was administered by telephone. To take under control potential carry-over effects, we randomly selected 158 participants that took the TICS first and were then administered the in-person screeners, with remaining participant ($N=310$) participants completing the in-person session before the TICS. The interval between the two sessions was approximately 5–7 days.

The TICS was administered in accordance with a previously established protocol [17]. Participants took the test while being at their place. A third person was required to witness to the administration of the TICS so that he/she could ensure that neither cues/facilitating items (e.g., a calendar; paper and pencil; etc.) nor distractions were present within the testing environment. This third party was also asked to confirm (or not) the correctness of examinees' responses to spatial orientation items. A brief sound-check was also conducted to verify that the quality of the audio was acceptable from both the examinee's and the examiner's standpoints.

In-person sessions followed this fixed administration sequence: MMSE, MoCA, and IP-TICS. To avoid redundant testing, overlapping items across these three tests (namely, the “Serial Sevens” task and some questions on spatial and temporal orientation) were only re-administered if the participant failed the preceding attempt. In alignment with a previously established procedure [17, 20], any item successfully completed in an earlier test was automatically scored as correct for all subsequent assessments.

Statistical analysis

Background analyses

Normality checks were run on each cognitive measure by addressing skewness values – judged as indexing a violation of the assumption if ≥ 1 [21] – and by visually inspecting histograms and Q-Q plots for abnormalities.

To explore whether the administration order – i.e., telephone-based first vs. in-person first – affected examinee's performances on the IP-TICS/IP-TICS_{&DR}, independent-sample *t*-tests were run on both IP-TICS and IP-TICS_{&DR} scores by addressing the abovementioned classification as the factor.

Convergent validity

Given that the vast majority of cognitive scores did not distribute normally, the convergent validity of the IP-TICS/IP-TICS_{&DR} against the TICS/TICS_{&DR} on one hand and the MMSE and MoCA on the other was tested *via* Spearman's coefficients. More specifically, correlations were run both by addressing total scores on these tests and subscale-level scores. The MMSE-O was left out of subscale-level correlational analyses given that the vast majority of its items overlapped with those of the IP-TICS-O. Whenever appropriate, Bonferroni's correction was applied to the significance threshold in order not to incur in an inflation of Type-I error rates – with α_{adjusted} being computed as $0.05/k$, with k being the number of correlations run. The magnitude of each Spearman coefficient was interpreted according to Cohen's [22] benchmarks – i.e., $r_s \leq 0.30 \rightarrow$ small; $0.30 < r_s \leq 0.50 \rightarrow$ medium; $r_s \geq 0.50 \rightarrow$ large. Following Hobart's [23] guidelines, which indicate that validity analyses should require at least 80 observations, the current sample size was considered as appropriate to this aim.

Measurement invariance

The measurement invariance between the TICS and IP-TICS was assessed by entering each sub-score – i.e., TICS-/IP-TICS-O, -L, -M and -AEF – in a hierarchical series of

Confirmatory Factor Analyses (CFA) for repeated measures [24] conducted by means of the *SEMLj* module in jamovi (<https://semlj.github.io/>) – which, in turn, implements the *lavaan* package in R [25]. The hierarchical procedure included three steps [24] – each based on the fact that the Italian TICS is underpinned by a mono-dimensional structure (“global cognition”) [14, 15]. First, we tested configural invariance, namely whether the same, mono-dimensional structure held for both the TICS and the IP-TICS. Subsequently, factor loadings were constrained to be equal across the two tests to assess whether each subscale contributed equally to the target construct regardless of the test being administered over the telephone vs. in person (metric invariance). Finally, sub-score intercepts were further constrained to be equal across the TICS and the IP-TICS to test scalar invariance – i.e., whether examinees’ performances, at the same level of the latent trait, are comparable across the two tests. Given that data were statistically dependent, residual covariances between corresponding sub-scores across the two tests were accounted for within each of these steps. Individual model fit was evaluated *via* the Comparative Fit Index (CFI), the Root Mean Square Error of Approximation (RMSEA) and the Standardized Root Mean Square Residual (SRMR) – being judged as appropriate if > 0.95 , ≤ 0.05 and ≤ 0.07 , respectively [26]. According to Chen [27], invariance was supported if the decrease in CFI values and the increase in RMSEA values between increasingly restrictive models was ≤ -0.01 and ≤ 0.015 , respectively. Hierarchical CFAs were also run to test the measurement invariance between the TICS_{&DR} and the IP-TICS_{&DR}. As reported by Kyriazos [28], at least 100 observations need to be considered to run a CFA; hence, our sample size was adequate to run these analyses.

Equivalence testing

The statistical equivalence between the TICS/TICS_{&DR} and the IP-TICS/IP-TICS_{&DR} was tested by means of Laken’s [29] two one-sided test (TOST) procedure. Such an approach allows to conclude on the statistical equivalence between two dependent means if the effect size of their difference falls within a prespecified equivalence interval (i.e., when the difference is simultaneously statistically greater than the lower bound and statistically smaller than the upper bound). TOST analyses were run via the dedicated jamovi macro relying on the R package *TOSTER* [29, 30]. Based on Lakens [29], a minimum of 44 pairs of observations are adequate for a TOST analysis when addressing a significance level α of 0.05, a power $1-\beta$ of 0.9, and equivalence bounds between -0.5 and 0.5 (these being default values in the absence of specific hypothesis regarding the smallest meaningful effect size).

Equating norms

Score equating was performed using the log-linear smoothed equipercentile equating (LSEE) approach [7] through the R package *equi* [31], producing conversion tables between TICS/TICS_{&DR} and IP-TICS/IP-TICS_{&DR}. The LSEE method converts measures by matching their percentile ranks with a log-linear function that reduces distribution irregularities [7]. To the aim of these analyses, our sample size included a sufficiently large number of observations based on Livingston’s [32] recommendations, which suggest a sample size of at least $N = 200$ for LSEE analyses.

In accordance with previous reports on the topic [17, 33, 34], a TOST procedure for dependent measures was subsequently to validate the estimated cross-walks, namely to determine whether equated scores derived through LSEE algorithms were statistically equivalent to empirical ones. In these analyses, LSEE-derived scores were rounded up in order for them to be integer given that empirical test scores do not foresee the presence of decimals.

Analyses were performed using R 4.3.0 (R Core Team, 2023), SPSS 29 (IBM Corp., 2023) and jamovi 2.6 (The jamovi project, 2024). Three participants did not complete the *Delayed Recall* task of the IP-TICS; therefore, analyses regarding this measure and the IP-TICS_{&DR} were run on 465 observations (vs. 468).

Results

Participants’ demographic and cognitive characteristics are summarized in Table 1. Participants being tested over the telephone first reported higher ($t(466) = -4.37$; $p < .001$) IP-TICS scores when compared to those undergoing the in-person screening session first (36.25 ± 1.98 vs. 35.30 ± 2.35). The same was true for the IP-TICS_{&DR} ($t(463) = -4.19$; $p < .001$; 39.77 ± 3.88 vs. 38.09 ± 4.17).

Convergent validity

The IP-TICS showed a large correlation with the TICS ($r_s(468) = 0.52$; $p < .001$), with the same being true for the association between the TICS_{&DR} and the IP-TICS_{&DR} ($r_s(468) = 0.56$; $p < .001$). Consistently, at $\alpha_{\text{adjusted}} = 0.025$, the IP-TICS and the IP-TICS_{&DR} showed medium-to-large association with both the MMSE (IP-TICS: $r_s(468) = 0.34$; $p < .001$; IP-TICS_{&DR}: $r_s(465) = 0.33$; $p < .001$) and the MoCA (IP-TICS: $r_s(468) = 0.48$; $p < .001$; IP-TICS_{&DR}: $r_s(465) = 0.53$; $p < .001$).

Table 2 shows the results of subscale-level convergent validity testing analyses between the TICS and the IP-TICS. Overall, at $\alpha_{\text{adjusted}} = 0.002$, each IP-TICS sub-scale showed

Table 1 Participants' demographic and cognitive data

N	468
Age (years)	57.5±15.4 (20–94)
Sex (male/female)	176/292
Education (years)	13.2±3.7 (5–25)
MMSE	29.3±0.9 (26–30)
MoCA	25.4±3 (14–30)
Executive function	3.3±0.9 (0–4)
Attention	5.5±0.9 (2–6)
Language	4.8±0.5 (2–5)
Visuo-spatial abilities	3.4±0.8 (0–4)
Orientation	5.9±0.2 (4–6)
Delayed Recall	2.5±1.7 (0–5)
TICS	35.9±2.4 (24–41)
Attention and Executive Functioning	8.6±0.9 (3–9)
Language	7.9±0.3 (6–8)
Memory	7.5±1.9 (0–12)
Orientation	11.9±0.3 (9–12)
Delayed Recall	4.1±2.2 (0–10)
TICS&DR	40±4.1 (24–51)
IP-TICS	35.6±2.3 (25–41)
Attention and Executive Functioning	8.8±0.8 (4–9)
Language	7.9±0.3 (6–8)
Memory	7.1±1.8 (1–12)
Orientation	11.9±0.4 (9–12)
Delayed Recall	3.1±2.2 (0–10)
IP-TICS&DR	38.7±4.1 (25–51)

IP-TICS=In-Person Telephone Interview for Cognitive Status; IPTICS&DR=In-Person Telephone Interview for Cognitive Status with Delayed Recall; MoCA= Montreal Cognitive Assessment; MMSE= Mini-Mental State Examination; TICS= Telephone Interview for Cognitive Status; TICS& DR= Telephone Interview for Cognitive Status with Delayed Recall. Continuous outcomes are shown as $M\pm SD$ (range), whilst categorical ones as frequencies and/or percentages

Table 2 Spearman coefficients for the association between TICS and IP-TICS subscales

	TICS-O	TICS-M	TICS-AEF	TICS-L	TICS-DR
IP-TICS-O	0.13	0.14	0.10	0.04	0.07
IP-TICS-M	0.14	.47*	0.16*	0.09	0.41*
IP-TICS-AEF	0.04	0.14	0.43*	0.05	0.14
IP-TICS-L	0.10	0.14	0.13	0.43*	0.17*
IP-TICS-DR	0.12	0.46*	0.13	0.05	0.53*

IP-TICS-O=In-Person Telephone Interview for Cognitive Status-Orientation; IP-TICS-M=In-Person Telephone Interview for Cognitive Status-Memory; IP-TICS-AEF=In-Person Telephone Interview for Cognitive Status-Attention and Executive Function; IP-TICS-L=In-Person Telephone Interview for Cognitive Status-Language; IP-TICS-DR=In-Person Telephone Interview for Cognitive Status-Delayed Recall; TICS-O= Telephone Interview for Cognitive Status-Orientation; TICS-M= Telephone Interview for Cognitive Status-Memory; TICS-AEF= Telephone Interview for Cognitive Status-Attention and Executive Function; TICS-L= Telephone Interview for Cognitive Status-Language; TICS-DR= Telephone Interview for Cognitive Status-Delayed Recall. *Significant coefficient at $\alpha_{\text{adjusted}} = 0.05/k = 0.05/25 = 0.002$

medium-to-large correlations with its TICS counterpart, with the exception of the *Orientation* one. In addition, and expectedly, the IP-TICS-M converged with the TICS-DR and, consistently, the IP-TICS-DR converged with the TICS-M. Two further association survived Bonferroni's correction – albeit being small in size: that between the IP-TICS-AEF and the TICS-M and that between the IP-TICS-L and the TICS-DR.

Subscale-level correlation coefficients between the IP-TICS and established in-person screeners (MMSE; MoCA) are displayed in Table 3 ($\alpha_{\text{adjusted}} = 0.001$). The IP-TICS-O expectedly showed a strong association with the MoCA-O only. The IP-TICS-M showed small-to-medium associations with the MMSE-M and all MoCA subscales except for the MoCA-O. The same pattern was detected as to the IP-TICS-DR – with the sole difference lying in the lack of association between this measure and the MoCA-VS. With regard to the IP-TICS-AEF, small-to-medium correlations were detected with the MMSE-A, the MMSE-L and MoCA-A. By contrast, no association involving the IP-TICS-L survived Bonferroni's correction.

Measurement invariance

The results of hierarchical CFAs aimed at assessing the measurement invariance between the TICS and the IP-TICS, both with and without the *Delayed Recall* task, are reported in Table 4. As to the TICS-vs.-IP-TICS comparison, the configural model supported a stable, single-factor structure across the two tests. Full metric invariance was also satisfied, with both individual fit indices and changes in CFI and RMSEA falling within acceptable ranges. However, scalar invariance happened to be rejected, with *post-hoc* constraint score tests revealing that this was driven by significant differences in the intercepts of the *Attention and Executive Functioning* ($\chi^2(1) = 32.42$; $p < .001$), *Memory* ($\chi^2(1) = 29.46$; $p < .001$) and *Language* ($\chi^2(1) = 4.17$; $p = .041$) subscales. When adding to the model the *Delayed Recall* task (i.e., TICS&DR-vs.-IP-TICS&DR comparison), the configural model still showed an adequate fit to the data, with full metric invariance being again supported as well. Moreover, and similarly to the first model, scalar invariance was not achieved – with *post-hoc* constraint score tests revealing that this was due to significant discrepancies in the intercepts of the *Delayed Recall* ($\chi^2(1) = 65.17$; $p < .001$) and *Attention and Executive Functioning* ($\chi^2(1) = 41.13$; $p < .001$) subscales.

Table 3 Spearman's coefficients for the association between IP-TICS and MMSE/MoCA subscales

	IP-TICS-O	IP-TICS-M	IP-TICS-AEF	IP-TICS-L	IP-TICS_DR
MMSE-M	0.06	0.20*	0.09	0.09	0.23*
MMSE-A	-0.04	0.10	0.24*	0.10	0.14
MMSE-L	0.02	0.13	0.18*	0.03	0.06
MoCA-VS	0.06	0.16*	0.10	0.14	0.14
MoCA-EF	0.09	0.25*	0.14	0.12	0.24*
MoCA-L	0.08	0.22*	0.14	0.11	0.18*
MoCA-A	0.08	0.33*	0.32*	0.11	0.31*
MoCA-DR	0.13	0.34*	0.14	0.10	0.43*
MoCA-O	0.65*	0.08	0.03	0.08	0.09

IP-TICS-O=In-Person Telephone Interview for Cognitive Status-Oriented; IP-TICS-M=In-Person Telephone Interview for Cognitive Status-Memory; IP-TICS-AEF=In-Person Telephone Interview for Cognitive Status-Attention and Executive Function; IP-TICS-L=In-Person Telephone Interview for Cognitive Status-Language; IP-TICS-DR=In-Person Telephone Interview for Cognitive Status-Delayed Recall; MMSE-M= Mini-Mental State Examination-Memory; MMSE-A= Mini-Mental State Examination-Attention; MMSE-L= Mini-Mental State Examination-Language; MoCA-VS= Montreal Cognitive Assessment-Visuo-Spatial abilities; MoCA-EF= Montreal Cognitive Assessment-Executive Function; MoCA-L= Montreal Cognitive Assessment-Language; MoCA-DR= Montreal Cognitive Assessment-Delayed Recall; MoCA-O= Montreal Cognitive Assessment-Oriented. *Significant coefficient at $\alpha_{\text{adjusted}} = 0.05/k = 0.05/45 = 0.001$

Equivalence testing

The results of the TOST procedure for testing the equivalence between the TICS/TICS_{&DR} and the IP-TICS/IP-TICS_{&DR} are shown in Table 5. In spite of a significant between-group difference as yielded by the dependent-sample *t*-test, the TICS proved to be equivalent to the IP-TICS – with both the lower and upper equivalence bounds falling within the predefined limits. At variance, equivalence between the TICS_{&DR} and

the IP-TICS_{&DR} was rejected – with scores on the former test being higher than those on the latter.

Equating norms

Table 6 reports LSEE-based crosswalks to convert TICS/TICS_{&DR} scores to those on the IP-TICS/IP-TICS_{&DR}, and vice-versa. Conversions could not be derived for TICS/TICS_{&DR} and IP-TICS/IP-TICS_{&DR} scores below 24 and 25, respectively, since such lowermost scores were not represented within the present sample. For the same reason, no counterparts were found for TICS/TICS_{&DR} scores equal to 25, 26 and 28, as well as for IP-TICS/IP-TICS_{&DR} scores equal to 26 and 27. Within TICS-to-IP-TICS and IP-TICS-to-TICS crosswalks, both the source and the equated score were mostly identical. Similarly, as to TICS_{&DR}-to-IP-TICS_{&DR} and IP-TICS_{&DR}-to-TICS_{&DR} crosswalks, the raw difference between source and equated scores were minimal and typically ranged from 0 to |2|.

The results of TOST analyses aimed at validating the aforementioned crosswalks are reported in Table 7. Equivalence between empirical and LSEE-derived scores was supported for each cross-walk.

Discussion

The present study provides the Italian validation of the in-person version of the TICS to be administered face-to-face (i.e., the IP-TICS), supporting its use in clinical and

Table 4 Results of hierarchical CFAs testing the measurement invariance between the TICS and the IP-TICS (both with and without the *Delayed Recall* task)

		χ^2	df	<i>p</i>	CFI	RMSEA [95% CI]	SRMR	Δ CFI	Δ RMSEA	Invariance
TICS	Configural	27.75	15	0.023	0.976	0.043 [0.016, 0.067]	0.031	-	-	Supported
	Metric	31.70	18	0.024	0.974	0.040 [0.015, 0.063]	0.034	0.002	0.003	Supported
	Scalar	101.33	22	<0.001	0.851	0.088 [0.071, 0.105]	0.053	0.123	0.048	Rejected
IP-TICS	Configural	68.74	29	<0.001	0.971	0.054 [0.038, 0.071]	0.049	-	-	Supported
	Metric	74.92	33	<0.001	0.969	0.052 [0.037, 0.068]	0.051	0.002	0.002	Supported
	Scalar	225.60	38	<0.001	0.863	0.103 [0.090, 0.116]	0.080	0.106	0.051	Rejected

CFA=Confirmatory Factor Analysis; TICS=Telephone Interview for Cognitive Status; IP-TICS=In-Person Telephone Interview for Cognitive Status; CFI=Comparative Fit Index; RMSEA=Root Mean Square Error of Approximation; SRMR=Standardized Root Mean Square Residual

Table 5 Results of the TOST procedure for testing the equivalence between TICS/TICS_{&DR} and IP-TICS/IP-TICS_{&DR}

TICS		IP-TICS		<i>t</i>	<i>p</i>	Equivalence
TICS	vs.	IP-TICS	<i>t</i> -test	3.3	<0.001	Supported
35.9±2.38		35.6±2.28	Upper EB	8.6	<0.001	
			Lower EB	1.8	0.033	
TICS _{&DR}	vs.	IP-TICS _{&DR}	<i>t</i> -test	-7.9	<0.001	Rejected
40±4.13		38.7±4.14	Upper EB	10.8	<0.001	
			Lower EB	-4.9	1.00	

IP-TICS=In-Person Telephone Interview for Cognitive Status; IP-TICS_{&DR}=In-Person Telephone Interview for Cognitive Status with Delayed Recall; TICS= Telephone Interview for Cognitive Status; TICS_{&DR}= Telephone Interview for Cognitive Status with Delayed Recall; EB=equivalent bound. Continuous outcomes are shown as *M*±*SD* (range)

Table 6 LSEE-derived tables to convert TICS/TICS_{&DR} scores to those on the IP-TICS/IP-TICS_{&DR}, and vice-versa

TICS to IP-TICS		IP-TICS to TICS		TICS _{&DR} to IP-TICS _{&DR}		IP-TICS _{&DR} to TICS _{&DR}	
TICS	IP-TICS	IP-TICS	TICS	TICS _{&DR}	IP-TICS _{&DR}	IP-TICS _{&DR}	TICS _{&DR}
24	25	25	25	24	26	25	24
25	-	26	-	25	-	26	-
26	-	27	-	26	-	27	-
27	26	28	28	27	27	28	28
28	-	29	29	28	-	29	30
29	29	30	30	29	28	30	31
30	30	31	31	30	29	31	32
31	31	32	32	31	30	32	34
32	32	33	33	32	30	33	35
33	33	34	34	33	31	34	36
34	34	35	35	34	32	35	36
35	35	36	36	35	33	36	37
36	36	37	37	36	34	37	38
37	37	38	39	37	36	38	39
38	38	39	40	38	37	39	40
39	39	40	41	39	38	40	41
40	40	41	41	40	39	41	42
41	41			41	40	42	43
				42	41	43	44
				43	42	44	45
				44	43	45	46
				45	44	46	48
				46	45	47	49
				47	46	48	50
				48	46	49	50
				49	47	51	51
				50	48		
				51	51		

IP-TICS=In-Person Telephone Interview for Cognitive Status; IP-TICS_{&DR}=In-Person Telephone Interview for Cognitive Status with Delayed Recall; TICS= Telephone Interview for Cognitive Status; TICS_{&DR}=Telephone Interview for Cognitive Status with Delayed Recall

research settings where assessment modality may vary or standard screening procedures are difficult to implement. More specifically, this study supports the convergence, equivalence and configural and metric invariance between the IP-TICS and the TICS, also providing evidence for the construct validity of the IP-TICS against two established in-person cognitive screening tests: the MMSE and the MoCA. In addition, we deliver LSEE-based, straightforward conversion table to estimate TICS scores from IP-TICS ones, and vice-versa.

Convergence, invariance and equivalence between the TICS and the IP-TICS

A clinically relevant question concerns whether the IP-TICS and the telephone-based TICS provide comparable measures of cognitive functioning, thus allowing clinicians and researchers to interpret scores consistently. We approached this matter by exploring the convergent validity, measurement invariance and statistical equivalence between the two measures. As to convergent validity analyses, the IP-TICS

proved to be strongly associated with the TICS – this being true also when including the *Delayed Recall* task. Notably, moderate-to-strong correlations were also found at the sub-scale level – with each IP-TICS sub-score, except for the IP-TICS-O, converging with its telephone-based counterparts. These findings mostly align with those yielding from measurement invariance testing. In fact, full configural and metric invariance between the TICS/TICS_{&DR} and the IP-TICS/IP-TICS_{&DR} was supported, this suggesting that: (1) the mono-dimensional structure of the TICS is stable regardless of whether it is administered over the telephone or in person and (2) each of its sub-scale contributes equally to the target construct (i.e., “global cognition”) across the two formats. Taken together, these findings avail the notion that the IP-TICS/IP-TICS_{&DR} is overall comparable to the TICS/TICS_{&DR} in terms of what is being measured – namely, global cognition/overall cognitive efficiency.

At the same time, full scalar invariance between the TICS and the IP-TICS (as well as between the TICS_{&DR} and the IP-TICS_{&DR}) was not met, suggesting that administration modality may influence performance on specific

Table 7 TOST procedure for the comparison between empirical and LSEE-derived scores

Empirical score		Derived score		t	p	Equivalence
TICS	vs.	TICS derived from IP-TICS	<i>t</i> -test	1.4	0.152	Supported
35.9±2.38		35.8±2.50	Upper EB	3.4	<0.001	
			Lower EB	6.3	<0.001	
IP-TICS	vs.	IP-TICS derived from TICS	<i>t</i> -test	3.3	0.001	Supported
35.6±2.28		35.9±2.40	Upper EB	8.5	<0.001	
			Lower EB	1.9	0.030	
TICS _{&DR}	vs.	TICS _{&DR} derived from IP-TICS _{&DR}	<i>t</i> -test	1.1	0.276	Supported
40±4.13		39.8±4.11	Upper EB	4.0	<0.001	
			Lower EB	1.9	0.032	
IP-TICS _{&DR}	vs.	IP-TICS _{&DR} derived from TICS _{&DR}	<i>t</i> -test	1.0	0.323	Supported
38.7±4.14		38.8±4.28	Upper EB	1.9	0.030	
			Lower EB	3.9	<0.001	

IP-TICS=In-Person Telephone Interview for Cognitive Status; IPTICS_{&DR}=In-Person Telephone Interview for Cognitive Status with Delayed Recall; TICS= Telephone Interview for Cognitive Status; TICS_{&DR}= Telephone Interview for Cognitive Status with Delayed Recall; EB=equivalent bound. Continuous outcomes are shown as $M \pm SD$ (range)

subdomains despite overall comparability between formats. When looking at *post-hoc* constraint score tests, *Attention and Executive Functioning*, *Memory* and – although marginally – *Language* subscales contributed to the lack of scalar invariance within the TICS-vs.-IP-TICS comparison, while this was true for the *Attention and Executive Functioning* and *Delayed Recall* subscales when comparing the TICS_{&DR} to the IP-TICS_{&DR}. Such findings hint at the fact that, at least when dealing with attentional-executive and memory sub-scales, construct-irrelevant confounders related to the different modality of administration account for potential discrepancies in test scores between the TICS/TICS_{&DR} and the IP-TICS/IP-TICS_{&DR}.

While seemingly contradictory, these findings are mostly in line with the results from the TOST procedure. Indeed, although the IP-TICS was found to be statistically equivalent to the TICS, this equivalence did not extend to the TICS_{&DR}-vs.-IP-TICS_{&DR} comparison. Furthermore, despite the effect size of the difference between IP-TICS and TICS scores falling within the predefined equivalence bounds, the dependent-sample *t*-test still revealed a significant discrepancy between the two measures. It might thus be the case that the significant *t*-test result at the total score level is driven by the subscale-level intercept differences identified by the scalar invariance CFA, especially within the *Attention and Executive Functioning* and *Memory* subscales. Taken together, these results suggest that, while subscale-level inconsistencies might exist between the IP-TICS and the TICS, such differences are likely to be attenuated when addressing the total score, thereby leading to the two tests being statistically equivalent at a global level. One possible interpretation is that modality-specific performance variations concerning a given subscale – e.g., *Memory* – may be counterbalanced by more stable scores on tasks

enclosed within domains that are less subject to variability – e.g., *Orientation* –, thereby preserving the equivalence of total scores across the two formats. From another perspective, it could be argued that such total-score equivalence might also reflect the masking of localized mode effects rather than absolute construct equivalence across all subdomains – a possibility that warrants further investigations.

Notably, this pattern did not extend to the versions of the test also including the *Delayed Recall*, suggesting that modality-related effects may become more evident when delayed memory performance is considered. In this last case, the subscale-level discrepancies are likely to be too large to be diluted within the total score. In the case of our cohort, participants happened to score higher on the *Delayed Recall* task when this was administered over the telephone. As previously suggested [5], it is possible that the telephone *medium* reduces performance anxiety related to highly demanding tasks – such as those requiring the delayed retrieval of verbal items [35] – due to the lack of physical examiner-examinee interactions and the fact that the examinee is tested in a familiar environment (i.e., his/her place).

It is interesting to note that most of the abovementioned findings closely align with a recent, large-scale study carried out by Smith et al. [5], where the comparability between telephone-based and in-person performances on some tasks drawn from the TICS was assessed in a population-based sample of 6,825 individuals aged ≥ 65 years. Although these Authors⁵ embraced an independent-sample study design (i.e., some individuals taking the tests over the telephone and some others in person), they likewise found that, despite cross-modality configural, metric and scalar invariance being met, significant discrepancies emerged in the intercepts of memory and attention measures.

Crosswalks between the IP-TICS and the TICS

A key contribution of our study lies in the provision of equating norms to convert IP-TICS/IP-TICS_{&DR} scores to TICS/TICS_{&DR} ones, and vice-versa, providing clinicians with a practical tool to compare cognitive performance across telephone and face-to-face assessments. Given the mixed results of measurement invariance analyses, which do not entirely support the full comparability between the two tests, the availability of such conversion tables plays an even more important role towards clinical and research practice. In fact, these crosswalks properly allow to level out the between-modality discrepancies in test scores that have been outlined by measurement invariance and equivalence analyses. Most relevantly, these conversions happened to be supported by TOST analyses comparing empirical to LSEE-derived scores – thus providing evidence for the validity of the crosswalks themselves.

As to TICS-to-IP-TICS and IP-TICS-to-TICS conversions, it is interesting to note that source and equated scores were often identical, with the exception of lowermost values. This occurrence suggests that the lack of scalar invariance reported above might selectively apply, at least to some extent, to those examinees scoring on the lower range of the test. This hypothesis is once again supported by Smith et al.'s [5] study, which demonstrated that discrepancies between the different modalities are particularly evident in examinees with poor performances.

With regard to the crosswalks for the test formats including the *Delayed Recall* task, source and target scores were only infrequently identical, in line with both the lack of equivalence and of scalar invariance between the TICS_{&DR} and the IP-TICS_{&DR}.

Construct validity of the IP-TICS against established in-person screeners

From a clinical perspective, establishing whether the IP-TICS captures cognitive constructs comparable to those assessed by widely used screening measures is essential to support its interpretability within routine assessment pathways. Thus, a further, relevant findings lie in the demonstration that the total score on the IP-TICS – both with and without the *Delayed Recall* task – moderately-to-strongly converges with widespread cognitive screening tests such as the MMSE and the MoCA. This suggests that the IP-TICS/IP-TICS_{&DR} captures cognitive constructs broadly comparable to established in-person screeners. This is also consistent with a recent Italian study [17] supporting the convergent validity between TICS/TICS_{&DR} scores and both the MMSE and the MoCA.

The results of correlational analyses at a total score level mostly aligned with those at the subscale level – with significant associations being found between IP-TICS sub-scores and MMSE/MoCA ones aimed at measuring similar cognitive domains/functions, with the exception of the IP-TICS-L. In fact, this last subscale showed not significant associations with MMSE/MoCA sub-scores after Bonferroni's correction. A possible explanation to this finding might lie in the relatively low variability of language measures across the three tests. This is consistent with the notion that ceiling effects are often found in healthy individuals' scores on simple language tests [27, 36]. A further finding that is worth discussing is represented by the fact that the IP-TICS-M and the IP-TICS-DR showed a widespread pattern of associations with MMSE/MoCA sub-scores. More specifically, these two IP-TICS sub-scores not only converged with MMSE/MoCA sub-scale assessing memory, but also with those assessing executive functioning, attention, language and, with specific regard to the IP-TICS-M, visuo-spatial skills. This finding is consistent with the notion that the performance on memory tasks enclosed within multi-component screening tests reflect an integrative process engaging multiple cognitive systems [35].

Implications for clinical practice and research

This study adds up to and complements the growing Italian literature on TBCS tests [10, 14, 15, 37–41], and, more specifically, on the matter of the comparability between TBCS tests and in-person cognitive screeners [10, 13, 17].

First, our report delivers evidence regarding the potential application of the TICS and the IP-TICS within both clinical practice and research settings. In fact, the convergence, configural/metric invariance and equivalence between the two tests warrant that the underlying construct(s) being measured is stable across telephone and in-person modalities. The association between the IP-TICS and established in-person screening tests (MMSE; MoCA) further strengthens its validity as measure of global cognition. The availability of the present crosswalks to convert IP-TICS scores to those on the TICS – and vice-versa – closes the circle, as allowing for a proper comparability between these two tests by overcoming the lack of scalar invariance. Based on these findings, clinicians and researchers may adopt more flexible cognitive screening pathways by adapting assessment modality according to patient characteristics and contextual constraints while preserving comparability of cognitive estimates across evaluations. In particular, compatibility between the TICS and IP-TICS may be especially relevant when standard screening procedures are not feasible (e.g., bedside evaluations, individuals with visual or

motor limitations, isolation settings, or longitudinal protocols alternating telephone and face-to-face assessments).

For instance, a patient can be screened at baseline *via* the IP-TICS and then followed-up with the TICS, and vice-versa, without the administration modality preventing users from properly comparing longitudinal scores. In addition, within research settings, our conversion tables could help harmonize data collected *via* different formats of the test – especially within longitudinal, multi-center investigations. Finally, as suggested by Ferrucci et al. [13], this report supports the use of the IP-TICS as a valid proxy of global cognitive functioning in contexts where standard in-person cognitive screening tests are not feasible, such as when visual, motor, or setting-related confounders interfere with the administration and/or execution of tasks requiring visual information processing (from both the examiner's and the examinee's standpoints) or the provision of motor responses [43]. By contrast, as to the TICS_{&DR}/IP-TICS_{&DR}, the lack of equivalence between the telephone-based and in-person formats – in spite of the other findings being comparable to those addressing the TICS/IP-TICS – suggests that extreme caution should be exerted when comparing the two tests if the *Delayed Recall* task is also administered. Hence, also in the light of the facts that, firstly, this last sub-score proved not to increase the degree of informativity toward examinees' cognitive *status* within the Italian standardization of the TICS [14], secondary, the Italian version of the TICS without the *Delayed Recall* task still showed optimal clinimetric properties for the detection of mild cognitive impairment and dementia [41], we advise clinicians and researchers to focus on the “original” TICS/IP-TICS total (i.e., the one ranging 1–41) unless it is strictly relevant to their goals to additionally consider the *Delayed Recall* sub-score.

Limitations and future perspective

This study is of course not devoid of limitations. First and foremost, our sample only included healthy individuals, this limiting the generalizability of our findings toward either cognitively unimpaired individuals or examinees with subjective/mild cognitive dysfunctions. Accordingly, these findings should be interpreted as preliminary evidence of psychometric comparability, while further validation in clinical sample is essential before adopting widespread utilization in diagnostic protocols. Such a stance is mostly evident with regard to the present equating norms – which are unlikely to be useful for patients with moderate-to-severe cognitive impairment, given that conversions for lower-most scores (i.e., ≤ 24 for the TICS/IP-TICS and ≤ 25 for the TICS_{&DR}/IP-TICS_{&DR}) could not be estimated by the LSEE. Hence, a key goal that should be addressed within future studies on the topic would be to replicate the present

analyses in patients with brain disorders [44]. Second, this study did not focus on providing normative data for the IP-TICS. In spite of the demonstrated equivalence, we caution against the straightforward application of the existing Italian normative framework of the TICS [15] to the IP-TICS in formal clinical settings. In fact, unknown sources of systematic error variance inherent to the face-to-face modality might still influence examinees' performances, thus potentially undermining the validity of norms delivered for a test to be administered over the telephone [45]. Nevertheless, one possibility is that clinicians may cautiously apply TICS norms to the IP-TICS for exploratory purposes, provided that the lack of IP-TICS-specific normative data is explicitly acknowledged. By contrast, for research purposes, such an application may be considered more acceptable – although, once again, with caution, meaning that the potential for modality-related measurement errors is carefully accounted for when interpreting the results. With that being said, in the light of these considerations, it would be advisable for future studies to aim at deriving Italian norms for the IP-TICS. A further issue lies in the fact that some practice effects have been detected in relation to the IP-TICS: participants who took the TICS first scored higher on the IP-TICS than those who took the latter test before the TICS. Consequently, the relatively short interval between administrations (5–7 days) may have introduced carry-over effects, potentially increasing score similarity across formats, inflating equivalence estimates and, to some extent, influencing the derived conversion norms. Similarly, associations between the IP-TICS and established in-person screening measures should be interpreted cautiously, as MMSE and MoCA were systematically administered before the IP-TICS, potentially introducing order-related influences or practice effects. Moreover, identical items overlapping across screening measures were not re-administered within the IP-TICS if previously administered. Therefore, some scores may have been influenced by earlier assessments, potentially introducing circularity and inflating associations with MMSE and MoCA. Therefore, convergent validity findings should be interpreted cautiously.

Conclusions

This study supports the convergence, equivalence and configural and metric invariance between the TICS and the IP-TICS in an Italian population sample, offering promising evidence of their comparability across telephone and face-to-face administrations. Moreover, the IP-TICS converges with established in-person cognitive screening tests, supporting its potential utility in clinical settings where flexible assessment pathways or alternatives to standard screeners may be needed.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10072-026-09201-2>.

Acknowledgements The Authors are thankful to participants. We thank Giunti Psychometric for allowing us to employ the Telephone Interview for Cognitive Status for experimental purposes. We thank Dr. Ziad Nasreddine and his team for allowing us to employ the Montreal Cognitive Assessment for experimental purposes. We thank Dr. Lorenzo Diana for his help in data collection.

Authors' contribution ENA: conceptualization, data collection, analyses, drafting, revision; EG: data collection, revision; GDL: data collection, data curation, drafting, revision; FF: drafting, revision; BC, AM: data collection, data curation, drafting, revision; LE, SC, CC, AZ, EC, MAS: data collection, revision; FV: revision; VS, NT, BP: resources, revision; NB: conceptualization, resources, revision.

Funding Open access funding provided by Università degli Studi di Milano - Bicocca within the CRUI-CARE Agreement. This work was supported by Italian Ministry of Health – Ricerca Corrente.

Data availability Datasets associated with the present study cannot be made publicly available on ethical-legal grounds but have been stored on an online repository with restricted access (<https://doi.org/10.5281/zenodo.21361399>) and can be made available upon reasonable request of interested researchers to the Corresponding Author(s), who will forward a request for a data transfer agreement to the relevant Ethical Committee. Data concerning the Telephone Interview for Cognitive Status cannot be made available due to copyright-related reasons.

Declarations

Ethical approval This study was approved by the Ethics Committee of IRCCS Istituto Auxologico Italiano (I.D.: 25C122, 23C308). This research was conducted ethically in accordance with the World Medical Association Declaration of Helsinki.

Consent to participate Participants provided written informed consent.

Consent for publication Not applicable.

Competing interests E.N.A. and N.B. have edited the official manual of the Italian version of the Telephone Interview for Cognitive Status (TICS) on behalf of Giunti Psychometrics and perceive royalties from Giunti Psychometrics when such a manual is purchased. Giunti Psychometrics holds the copyright of the Italian TICS.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Castanho TC, Amorim L, Zihl J, Palha JA, Sousa N, Santos NC (2014) Telephone-based screening tools for mild cognitive impairment and dementia in aging studies: a review of validated instruments. *Front Aging Neurosci* 6:16. <https://doi.org/10.3389/fnagi.2014.00016>
- Elliott E, Green C, Llewellyn DJ, Quinn TJ (2020) Accuracy of telephone-based cognitive screening tests: systematic review and meta-analysis. *Curr Alzheimer Res* 17:460–471. <https://doi.org/10.2174/1567205017999200626201121>
- Herr M, Ankri J (2013) A critical review of the use of telephone tests to identify cognitive impairment in epidemiology and clinical research. *J Telemed Telecare* 19(1):45–54. <https://doi.org/10.1177/1357633X12474962>
- Booth T, Murray A, Muniz-Terrera G (2021) Are we measuring the same thing? Psychometric and research considerations when adopting new testing modes in the time of COVID-19. *Alzheimer's Dement* 17:251–254. <https://doi.org/10.1002/alz.12197>
- Smith JR, Gibbons LE, Crane PK, Mungas DM, Glymour MM, Manly JJ (2023) Shifting of cognitive assessments between face-to-face and telephone administration: measurement considerations. *J Gerontol B Psychol Sci Soc Sci* 78:191–200. <https://doi.org/10.1093/geronb/gbac135>
- Hlavka JP, Kinoshita AT, Fang S, Hunt A (2021) Clinical outcome measure crosswalks in Alzheimer's disease: a systematic review. *J Alzheimers Dis* 83:591–608. <https://doi.org/10.3233/JAD-210060>
- Kolen MJ, Brennan RL (2013) Test equating: Methods and practices. Springer Science & Business Media
- Larner AJ (2021) Cognitive testing in the COVID-19 era: can existing screeners be adapted for telephone use? *Neurodegener Dis Manag* 11(1):77–82. <https://doi.org/10.2217/nmt-2020-0040>
- Mondini S, Montemurro S, Pucci V, Ravelli A, Signorini M, Arcara G (2022) Global examination of mental state: an open tool for the brief evaluation of cognition. *Brain Behav* 12:e2710. <https://doi.org/10.1002/brb3.2710>
- Montemurro S, Mondini S, Pucci V, Durante G, Riccardi A, Maffezzini S et al (2023) Tele-Global examination of mental state (Tele-GEMS): an open tool for the remote neuropsychological screening. *Neurol Sci* 44:3499–508. <https://doi.org/10.1007/s1072-023-06862-1>
- Brandt J, Spencer M, Folstein M (1988) The telephone interview for cognitive status. *Neuropsychiatry Neuropsychol Behav Neurol* 1:111–117
- Folstein M, Folstein S, McHugh P (1975) Mini-mental state: a practical method for grading the cognitive state of patients for the clinician. *J Psychiatry Res* 12:189–198
- Ferrucci L, Del Lungo I, Guralnik JM, Bandinelli S, Benvenuti E, Salani B et al (1998) Is the telephone interview for cognitive status a valid alternative in persons who cannot be evaluated by the Mini Mental State Examination? *Aging Clin Exp Res* 10:332–8. <https://doi.org/10.1007/BF03339796>
- Aiello EN, Esposito A, Giannone I, Diana L, Appollonio I, Bolognini N (2022) Telephone Interview for Cognitive Status (TICS): Italian adaptation, psychometrics and diagnostics. *Neurol Sci* 43:3071–3077. <https://doi.org/10.1007/s10072-021-05729-7>
- Aiello EN, Diana L, Bolognini N, Appollonio I, Ferrucci L (2022) Telephone Interview For Cognitive Status™: Manuale di somministrazione e scoring. Giunti Psychometrics
- Nasreddine ZS, Phillips NA, Bédirian V, Charbonneau S, Whitehead V, Collin I et al (2005) The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *J Am Geriatr Soc* 53:695–9. <https://doi.org/10.1111/j.1532-5415.2005.53221.x>
- Aiello EN, Curti B, De Luca G, Casartelli S, Esposti L, Curatoli C et al (2025) Convergence and equating norms between the Telephone Interview for Cognitive Status (TICS), the MMSE and the MoCA in an Italian population sample. *Aging Clin Exp Res* 37:154. <https://doi.org/10.1007/s40520-025-03026-3>
- Foderaro G, Isella V, Mazzone A, Biglia E, Di Gangi M, Pasotti F et al (2022) Brand new norms for a good old test: Northern Italy

- normative study of MiniMental State Examination. *Neurol Sci* 43:3053–63. <https://doi.org/10.1007/s10072-021-05845-4>
19. Aiello EN, Gramegna C, Esposito A, Gazzaniga V, Zago S, Difonzo T et al (2021) The Montreal Cognitive Assessment (MoCA): updated norms and psychometric insights into adaptive testing from healthy individuals in Northern Italy. *Aging Clin Exp Res* 34(2):375–382. <https://doi.org/10.1007/s40520-021-01943-7>
 20. Dubbelman MA, Van De Beek M, Van Gils AM, Leeuwis AE, Van Der Vlies AE, Pijnenburg YA et al (2024) Comparing and linking the Mini-Mental state examination and Montreal cognitive assessment in the Amsterdam dementia cohort. *J Int Neuropsychol Soc* 30:867–74. <https://doi.org/10.1017/S1355617724000341>
 21. Kim HY (2013) Statistical notes for clinical researchers: assessing normal distribution (2) using skewness and kurtosis. *Restor Dent Endod* 38:52–4. <https://doi.org/10.5395/rde.2013.38.1.52>
 22. Cohen J (1992) Quantitative methods in psychology: a power primer. *Psychol Bull* 112:1155–9
 23. Hobart JC, Cano SJ, Warner TT, Thompson AJ (2012) What sample sizes for reliability and validity studies in neurology? *J Neurol* 259:2681–94. <https://doi.org/10.1007/s00415-012-6570-y>
 24. Vandenberg RJ, Lance CE (2000) A review and synthesis of the measurement invariance literature: suggestions, practices, and recommendations for organizational research. *Organ Res Methods* 3:4–70. <https://doi.org/10.1177/109442810031002>
 25. Rosseel Y (2012) Lavaan: an R package for structural equation modeling. *J Stat Softw* 48:1–36. <https://doi.org/10.18637/jss.v048.i02>
 26. West SG, Taylor AB, Wu W (2012) Model fit and model selection in structural equation modeling. In: Hoyle RH (ed) *Handbook of structural equation modeling*. Guilford Press, New York, pp 209–31
 27. Chen FF (2007) Sensitivity of goodness of fit indexes to lack of measurement invariance. *Struct Equ Model* 14:464–504. <https://doi.org/10.1080/10705510701301834>
 28. Kyriazos TA (2018) Applied psychometrics: sample size and sample power considerations in factor analysis (EFA, CFA) and SEM in general. *Psychology* 9:2207. <https://doi.org/10.4236/psych.2018.98126>
 29. Lakens D (2017) Equivalence tests. *Soc Psychol Personal Sci* 8:355–362. <https://doi.org/10.1177/1948550617697177>
 30. Caldwell AR (2022) Exploring Equivalence Testing with the Updated TOSTER R Package [Internet]. *PsyArXiv*; Available from: osf.io/preprints/psyarxiv/ty8de
 31. Wołodźko T, Kondrątek B, Szaleniec H (2014) Zrównanie ekwicyentylowe na tle innych metod zrównywania na przykładzie sprawdzianu i egzaminu gimnazjalnego. *Edukacja* 3:29
 32. Livingston SA (1993) Small-sample equating with log-linear smoothing. *J Educ Meas* 30:23–39. <https://doi.org/10.1111/j.1745-3984.1993.tb00420.x>
 33. Aiello EN, Pasotti F, Appollonio I, Bolognini N (2022) Equating mini-mental state examination (MMSE) and Montreal cognitive assessment (MoCA) scores: conversion norms from a healthy Italian population sample. *Aging Clin Exp Res* 34:1721–4. <https://doi.org/10.1007/s40520-022-02089-w>
 34. Aiello EN, Solca F, Greco LC, Torre S, Carelli L, Morelli C et al (2023) Equating norms between the ALS Cognitive Behavioral Screen (ALS-CBS™) and the Edinburgh Cognitive and Behavioural ALS Screen (ECAS) in non-demented ALS patients. *J Neurol* 270:4090–5. <https://doi.org/10.1007/s00415-023-11749-4>
 35. Dobbins IG, Foley H, Schacter DL, Wagner AD (2002) Executive control during episodic retrieval. *Neuron* 35(5):989–996. [https://doi.org/10.1016/S0896-6273\(02\)00858-9](https://doi.org/10.1016/S0896-6273(02)00858-9)
 36. Aiello EN, Pucci V, Diana L, Corvaglia A, Niang A, Mattiello S et al (2024) The Telephone Language Screener (TLS): standardization of a novel telephone-based screening test for language impairment. *Neurol Sci* 45:1989–2001. <https://doi.org/10.1007/s10072-023-07149-1>
 37. Aiello EN, Curti B, De Luca G, Moreschi A, Crispiatico V, Maranzano A et al (2025) Telephone-based cognitive screening in neurodegenerative MCI and dementia: preliminary findings from the TBCS Study. *Neurol Sci* 46:6509–19. <https://doi.org/10.1007/s10072-025-08580-2>
 38. Zanin E, Aiello EN, Diana L, Fusi G, Bonato M, Niang A et al (2022) Tele-neuropsychological assessment tools in Italy: a systematic review on psychometric properties and usability. *Neurol Sci* 43:125–138. <https://doi.org/10.1007/s10072-021-05719-9>
 39. Aiello EN, Esposito A, Giannone I, Diana L, Woolley S, Murphy J et al (2022) ALS Cognitive Behavioral Screen-Phone Version (ALS-CBS™-PhV): norms, psychometrics, and diagnostics in an Italian population sample. *Neurol Sci* 43:2571–2578. <https://doi.org/10.1007/s10072-021-05636-x>
 40. Trapp W, Röder S, Heid A, Billman P, Daiber S, Hajak G (2021) Sensitivity and specificity of the Bamberg Dementia Screening Test's (BDST) full and short versions: brief screening instruments for geriatric patients that are suitable for infectious environments. *BMC Med* 19(1):65. <https://doi.org/10.1186/s12916-021-01927-4>
 41. Bernini S, Ballante E, Picascia M, Barbieri M, Costa A, Cavallini E, Tassorelli C, Vecchi T, Bottiroli S (2025) Equating conversion norms for the Mini-Mental State Examination and Montreal Cognitive Assessment in healthy subjects and patients with neurodegenerative disorders. *Exp Gerontol* 205:112756. <https://doi.org/10.1016/j.exger.2025.112756>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Edoardo Nicolò Aiello¹  · Elena Guidetti² · Giulia De Luca¹ · Francesca Frisco¹ · Beatrice Curti¹ · Arianna Moreschi¹ · Lorenzo Esposti² · Sara Casartelli² · Chiara Curatoli¹ · Alice Zanin¹ · Elisa Camporeale¹ · Martina Andrea Sirtori^{2,3} · Federico Verde¹ · Vincenzo Silani^{1,4} · Nicola Ticozzi^{1,4} · Barbara Poletti^{1,5} · Nadia Bolognini^{2,6} 

✉ Nadia Bolognini
nadia.bolognini@unimib.it

Edoardo Nicolò Aiello
e.aiello@auxologico.it

Elena Guidetti
e.guidetti4@campus.unimib.it

Giulia De Luca
g.deluca@auxologico.it

Francesca Frisco
f.frisco@auxologico.it

Beatrice Curti
b.curti@auxologico.it

Arianna Moreschi
a.moreschi@auxologico.it

Lorenzo Esposti
l.esposti@campus.unimib.it

Sara Casartelli
sara.casartelli@unimib.it

Chiara Curatoli
chiara.curatoli@gmail.com

Alice Zanin
alice14297@hotmail.it

Elisa Camporeale
elisa.camporeale7@gmail.com

Martina Andrea Sirtori
m.sirtori@auxologico.it

Federico Verde
f.verde@auxologico.it

Vincenzo Silani
vincenzo@silani.com

Nicola Ticozzi
n.ticozzi@auxologico.it

Barbara Poletti
b.poletti@auxologico.it

¹ Department of Neurology and Laboratory of Neuroscience, IRCCS Istituto Auxologico Italiano, Milano, Italy

² Department of Psychology, University of Milano-Bicocca, Milano, Italy

³ Department of Neurosurgery, IRCCS Humanitas Research Hospital, Rozzano, Milan, Italy

⁴ Department of Pathophysiology and Transplantation, "Dino Ferrari" Center, Università degli Studi di Milano, Milano, Italy

⁵ Department of Oncology and Hemato-Oncology, Università degli Studi di Milano, Milano, Italy

⁶ Laboratory of Neuropsychology, Department of Neurorehabilitation Sciences, IRCCS Istituto Auxologico Italiano, Milano, Italy