JIOBA
Journal of Industrial and Organisational Behaviour in Africa
A peer-reviewed journal of the Nigerian Association of Industrial and Organisational Psychologists (NAIOP) • naiop.org.ng

e-ISSN: 0189-4285
https://doi.org/10.68422/jioba.jioba.v2i2.001

Vol. 2 • No. 2 • October 2026
pp. 1–13

RESEARCH ARTICLE PEER REVIEWED

AI-Assisted Adaptation of the Experiences of Social Inclusion Scale into Nigerian Pidgin: A Three-Wave Psychometric Study

Richards Ebireonwu Ebeh[1, *], Ngozi Nwakaego Sydney-Agbor[1], Olubukola Ajayi[2], Adanma Justina Solomon Nwagwu[3], Onyinyechi Mary Okorie[4], Mezewo Emerinwe Obuzor[5], Chijioke Obianuju Ahamefula-Aguguam[6] & Chukwuemeka Joseph Amadi[7]

[1] *Department of Psychology, Imo State University, Owerri, Nigeria*
[2] *Department of Psychology and Behavioural Studies, Ekiti State University, Ado Ekiti, Nigeria*
[3] *Department of Nursing, Imo State University, Owerri, Nigeria*
[4] *Department of Public Health, Federal University of Technology, Owerri, Nigeria*
[5] *Department of Social Work, Rivers State University, Nkpolu Oroworokwo, Port Harcourt, Nigeria*
[6] *General Studies Unit, University of Agriculture and Environmental Sciences, Umuagwo, Imo State, Nigeria*
[7] *Department of Psychology, Kingsley Ozumba Mbadiwe University, Ideato, Imo State, Nigeria*
*** Correspondence:** richebeh@gmail.com

RECEIVED	REVISED	ACCEPTED	PUBLISHED
31 May 2026	3 August 2026	6 September 2026	15 October 2026

ABSTRACT

Nigerian Pidgin serves as the everyday language, with estimates suggesting that between 75 and 100 million people use it (Faraclas, 1996; Eberhard et al., 2024), yet few psychometric instruments are available in this variety. The present study translated the 10-item Experiences of Social Inclusion Scale (ESIS; Leemann et al., 2022) into Nigerian Pidgin using an AI-assisted translation process. A within-subjects, three-wave repeated-measures design was used, in which 65 fourth-year psychology students at Imo State University, Owerri (M = 24.37 years, SD = 3.81; 24 men, 41 women) completed the original English (T1), Pidgin English (T2), and back-translated English (T3) versions at two-week intervals. Analyses included item-level descriptives, Cronbach's alpha, corrected item-total correlations and cross-version Pearson correlations. Internal consistency was acceptable to excellent across all three versions (T1 α = .81; T2 α = .93; T3 α = .90). Structural validity obtained through exploratory factor analysis of the Pidgin responses supported a single factor (KMO = .83; 60.96% of variance explained; loadings .44 to .97), with an average variance extracted of .61 and composite reliability of .94. Total scores on the three versions did not differ significantly, and every pair met statistical equivalence criteria at the ±0.5 pooled SD boundary. Four items displayed small but significant mean shifts between the original and back-translated versions, suggesting specific sites for revision. The study concludes that the Nigerian Pidgin version of the ESIS appears adaptable at this stage but would require further work with larger independent samples and confirmatory factor analysis before routine application.

KEYWORDS Nigerian Pidgin • Experiences of Social Inclusion Scale • AI-assisted adaptation • cross-cultural adaptation • emerging adults

HOW TO CITE THIS ARTICLE • APA 7
Ebeh, R. E., Sydney-Agbor, N. N., Ajayi, O., Nwagwu, A. J. S., Okorie, O. M., Obuzor, M. E., Ahamefula-Aguguam, C. O., & Amadi, C. J. (2026). AI-assisted adaptation of the Experiences of Social Inclusion Scale into Nigerian Pidgin: A three-wave psychometric study. *Journal of Industrial and Organisational Behaviour in Africa, 2*(2), 1–13. https://doi.org/10.68422/jioba.jioba.v2i2.001

Copyright: © 2026 Nigerian Association of Industrial and Organisational Psychologists (NAIOP). All rights reserved. No part of this article may be reproduced without the written permission of the publisher. Views expressed are those of the authors and do not necessarily reflect the position of NAIOP.

1. INTRODUCTION

Nigeria's linguistic ecology presents real problems for anyone producing measurement tools. Although English remains the language of government, courts, and tertiary education, ordinary life plays out in different registers altogether. In everyday life, many Nigerians conduct domestic, commercial, and informal social interactions in indigenous languages or Nigerian Pidgin, an English-lexifier creole that emerged through prolonged contact between English and West African languages during Atlantic trade and colonial expansion (Faraclas, 1996; Elugbe & Omamor, 1991; Deuber, 2005). Speaker estimates vary substantially depending on whether one counts first-language users, habitual users, or occasional users. Yakpo (2024) estimates the daily-user population at roughly 75 million, with some counts exceeding 100 million, which would rank Nigerian Pidgin among the most widely spoken languages in sub-Saharan Africa.

Despite this reach, Pidgin has remained at the margins of formal psychological measurement in Nigeria. Most adapted psychometric tools produced in the country over the past two decades have been rendered in Yoruba, Hausa, or Igbo, or simply left in English, with a thinner literature on Pidgin versions. Available translations include Yoruba renderings of the SF-12 (Mbada et al., 2021), an Igbo adaptation of the Geriatric Depression Scale (Mgbeojedo et al., 2022), the Yoruba EQ-5D (Fatoye et al., 2022), a Hausa MSPSS (Hamza et al., 2013), and a visual analogue scale adapted across the three major Nigerian languages (Odole & Akinpelu, 2009). These are meaningful steps, though they leave out the Pidgin-speaking population whose daily experience is captured in neither standard English nor in the three main regional languages. Respondents asked to complete an English instrument they understand imperfectly, or a regional-language instrument that does not map onto their own language practice, may struggle with item meaning and produce response patterns that carry more noise than substantive variance (Kaiser et al., 2022).

Social inclusion is a particularly relevant construct to bring into this conversation. Unlike classical outcome measures such as anxiety or depression, social inclusion captures the subjective side of belonging, meaningful activity, and perceived capacity to influence one's environment. The concept owes much to the capability approach developed by Sen (1999) and extended by Nussbaum (2011), which frames well-being not in terms of resources held but in terms of substantive opportunities to act and participate. From a capability lens, inclusion is less a fixed status than a bundle of functionings: feeling one's daily activities carry meaning, believing one matters to others, and judging that one has some realistic control over the direction of one's life. Leemann et al. (2022) built the Experiences of Social Inclusion Scale (ESIS) around this reading of inclusion, and the original Finnish validation returned a one-factor structure with Cronbach's alpha of .89 in a large population sample.

Emerging adults in Nigeria, by and large, occupy a life stage where this construct matters a great deal. Arnett's (2000, 2015) framing of the late teens to twenties as a distinct developmental period marked by identity exploration and instability has been cautiously extended to African contexts (Obidoa et al., 2019; Rotzinger et al., 2025; Thalmayer et al., 2024). Nigerian emerging adults confront intense labour market precarity, delayed markers of adulthood, and a well-documented mental health treatment gap (Fadele et al., 2024). Whether these young people feel included in the communities around them and whether they can name that experience are not trivial questions. The problem is that the instrument most directly designed to capture such experiences is available only in European languages, and whether its items travel well into Nigerian Pidgin remains untested.

There is an emerging body of work suggesting that creole-language adaptations can be done with acceptable psychometric consequences. Chukwu et al. (2023) translated the Rosenberg Self-Esteem Scale into Nigerian Pidgin and reported reasonable internal consistency among persons affected by leprosy. Murphy-Okpala et al. (2025) later extended this with Pidgin adaptations of stigma-related instruments in southern Nigerian adults. Kaiser and colleagues (2022) ran simultaneous validations of mental health screeners in Hausa and Pidgin and found that multilingual validation can surface linguistic biases easily missed in single-language work. Chukwuorji and Osondu (2023) mapped the overall literature on well-being assessment in African contexts and noted, with some concern, that Pidgin remains systematically under-represented.

A further wrinkle sits in the translation process itself. The conventional practice pairs forward and back-translation with bilingual human translators and expert committee review (Beaton et al., 2000; Wild et al., 2005). For Nigerian Pidgin, which lacks a widely agreed standard orthography among linguists and has no pool of certified translators with psychometric training, this practice is difficult to meet. Recent work on large language models has begun to probe whether AI-assisted translation can serve as a working substitute. Lu et al. (2025) benchmarked GPT-4, GPT-3.5, and Google Translate against professional translators on patient-reported outcome measures across six languages and reported largely comparable quality. Sørensen et al. (2025) ran a blinded non-inferiority trial comparing ChatGPT-4 and human translators on an abdominal

hernia outcome questionnaire. Grobelny et al. (2025) pushed the logic further by showing that an LLM-adapted organisational commitment scale could produce factor structure and measurement invariance properties broadly comparable to the human-adapted version. None of these studies, however, worked in a low-resource creole. The cautious note struck by AlKhamissi et al. (2024) on cultural alignment of LLMs is worth recalling: model outputs tend to align more closely with dominant-language cultural patterns, which may or may not generalise to Nigerian Pidgin.

Against this backdrop, the present study pilots an AI-assisted Pidgin adaptation of the ESIS. The empirical question is not whether the instrument is fully validated for routine use, a claim we are not in a position to make with 65 participants, but whether an AI-built Pidgin version carries enough of the original construct across a three-wave within-subjects test to justify further development work.

Purpose of the Study

The primary aim of this study is to translate and psychometrically evaluate the 10-item Experiences of Social Inclusion Scale (Leemann et al., 2022) in Nigerian Pidgin using AI-assisted translation. Specifically, the study sought to:

1. translate and culturally adapt the ESIS into Nigerian Pidgin using an AI-assisted process;
2. examine the internal consistency reliability of the Pidgin ESIS, including item-total statistics;
3. determine cross-version score equivalence between original English, Pidgin English,
4. back-translated English versions, and assess item-level back-translation fidelity to detect items that require additional editing; and
5. examine the structural validity of the Pidgin version through exploratory factor analysis of the Time 2 responses.

Hypotheses

H1: The Nigerian Pidgin version of the ESIS will demonstrate acceptable internal consistency (Cronbach's $\alpha \geq .70$).

H2: Total scores on the Pidgin English and back-translated English versions of the ESIS will not differ significantly from those on the original English version.

H3: There will be a strong positive correlation between the back-translated English version and the original English version ($r \geq .70$) of the ESIS.

H4: Comparison of the original and back-translated versions of the ESIS at the item level will show only a few semantically shifted items.

H5: The Nigerian Pidgin version of the ESIS will replicate the unidimensional structure reported for the original ESIS, with all items loading $\geq .40$ on a single retained factor.

2. EMPIRICAL REVIEW

The systematic cross-cultural adaptation of self-report measures is now a well-established practice, though one whose methodological standards have shifted subtly over time. Guillemin et al. (1993) and the subsequent elaboration by Beaton et al. (2000) established forward translation, back-translation, expert committee review, and pre-testing as the minimum defensible workflow. Wild and colleagues (2005), writing for the ISPOR Task Force, recommended multiple parallel forward translations, reconciliation, independent back-translation, expert harmonisation, cognitive debriefing, and a final proof-reading stage. More recent methodological scholarship has added layers around measurement invariance testing, differential item functioning, and cross-cultural structural validity assessment (Mokkink et al., 2010). The move has been from adequacy of translation toward adequacy of measurement across language versions, which is a subtly different question.

African adaptation work illustrates both the potential and the limits of these procedures. Fatoye et al. (2022) reported sound psychometric performance for a Yoruba EQ-5D among patients with musculoskeletal disorders. Mgbeojedo et al. (2022) documented similar results for an Igbo Geriatric Depression Scale. Igwesi-Chidobe et al. (2021) validated an Igbo MSPSS among Nigerians with chronic low back pain. Taken together, these studies suggest that with careful process, measurement in Nigerian indigenous languages can be reliably operationalised. Whether the same holds for Pidgin is a separate empirical question, and one that has only begun to receive attention.

Two recent Pidgin studies are particularly relevant here. Chukwu et al. (2023) reported that a Pidgin Rosenberg Self-Esteem Scale retained acceptable internal consistency among people affected by leprosy in southern Nigeria. Murphy-Okpala et al. (2025) extended this with Pidgin versions of stigma-related scales, again with defensible reliability and construct performance. These studies lend credibility to Pidgin as a target language for psychometric work, though both focused on populations experiencing specific health conditions rather than general emerging adults, and neither addressed positively framed well-being constructs.

Work on social inclusion measurement in African contexts is thinner still. Khumalo and colleagues (2010, 2013) constructed the General Psychological Well-Being Scale in a Setswana-speaking sample, and demonstrated that African conceptions of well-being can be operationalised in ways that are psychometrically sound. Bartolomei et al. (2024) reviewed how the capability approach has been translated into empirical well-being measurement more broadly, and their analysis suggests that capability-grounded instruments such as the ESIS are particularly suitable for contexts where material indicators alone miss key aspects of experience. Against this, Counted et al. (2024) emphasise that flourishing in African settings often sits inside communal frames whose Western scale analogues may or may not capture.

The question of AI-assisted adaptation brings a different literature into the conversation. Wang et al. (2024) showed that cross-lingual knowledge editing in large language models is effective for high-resource language pairs but substantially less so for low-resource varieties. Li et al. (2026) reported modest improvements for low-resource language adaptation in LLMs, though performance still lags behind output quality in dominant languages. Lu et al. (2025) compared LLM-translated PROMs against human translations across six languages and found substantial agreement. Sørensen et al. (2025) used a randomised non-inferiority design to show that ChatGPT-4 translations of an outcome questionnaire were not judged inferior to human translations. These studies, however, were limited to relatively well-resourced target languages. The extension to Nigerian Pidgin represents a step into territory which is, at best, thinly mapped.

The present study takes up that task. It offers a first empirical test of whether an AI-built Pidgin ESIS produces total scores and item-level responses that bear an acceptable relationship to the original instrument across three waves of administration, in a population for whom the ESIS is conceptually relevant.

3. METHOD

Research Design

A within-subjects three-wave repeated-measures design was used. Participants completed three instances of the ESIS at two-week intervals: the original English version at Time 1 (T1), the Nigerian Pidgin version at Time 2 (T2), and the back-translated English version at Time 3 (T3). Repeated administration to the same respondents offers the analytical advantage of removing between-person differences in social inclusion from the cross-version comparisons, so that each respondent serves as their own comparison unit. The design has a known trade-off: version effects are partially confounded with the passage of time and with response carryover. Social inclusion, however, is understood as a relatively stable subjective state across short intervals (Leemann et al., 2022), and a two-week gap was judged likely to attenuate verbatim memory of responses without allowing the construct itself to shift meaningfully.

Participants

A convenience sample of 65 undergraduates was drawn from the fourth-year cohort of the Department of Psychology, Imo State University, Owerri. Ages ranged from 20 to 45 years ($M = 24.37$, $SD = 3.81$), with 24 men and 41 women. Pidgin is a routine part of campus life in this setting, which made the sample appropriate for piloting the adaptation. All 65 participants completed all three waves of data collection, and no item-level missing data were recorded in any version.

Measure

The Experiences of Social Inclusion Scale (ESIS; Leemann et al., 2022) is a 10-item self-report instrument developed at the Finnish Institute for Health and Welfare. Items are rated on a 5-point Likert response format ranging from 1 (fully disagree) to 5 (strongly agree), yielding total scores between 10 and 50. Sample items include “I belong to a group or community that is important for me” and “I can strive for things that are important for me”. The scale was constructed around the capability approach of Sen (1999) and Nussbaum (2011) and is intended to measure the subjective experience of inclusion rather than its structural determinants. In the Finnish validation, internal consistency was $\alpha = .89$, factor analyses supported a single-factor solution, and convergent correlations with related constructs ranged from .41 to .68 (Leemann et al., 2022).

Translation Procedure

Forward and back-translation steps were conducted using artificial-intelligence applications rather than the conventional two-translator workflow. This choice was made under clear constraints. Nigerian Pidgin is a low-resource language for psychometric purposes, with neither an established pool of certified psychometric translators nor a widely accepted orthography. Recent work indicates that large language models can render meaningful cross-lingual content for underrepresented varieties, though quality tends to lag behind high-resource language pairs (Li et al., 2026; Wang et al., 2024). The original ESIS was translated into Nigerian Pidgin using Gemini (Google, 2026), a general-purpose large language model with documented multilingual coverage. The Pidgin version was then back-translated into English using Oversabi AI (Oversabi, 2026), an application trained with a focus on Nigerian language data, which was judged better suited to Pidgin lexical and syntactic features than a general-purpose model. We acknowledge that this procedure departs from the gold-standard practice outlined by Beaton et al. (2000) and Wild et al. (2005), and we consider the implications for translation fidelity in the limitations section.

Statistical Analyses

All analyses were conducted in Python 3, using the SciPy and statsmodels libraries, with reliability analyses cross-checked against SPSS output. Descriptive statistics, Cronbach's alpha, corrected item-total correlations, and cross-version Pearson correlations with 95% confidence intervals (via Fisher's r-to-z transformation) were computed for each version. Score interchangeability was estimated with the intraclass correlation coefficient ICC(2,1) and interpreted using Koo and Li's (2016) benchmarks. A one-way repeated-measures analysis of variance (RM-ANOVA) was used to test mean differences across versions, with partial eta-squared as the effect size and Bonferroni-corrected paired *t*-tests for follow-up contrasts. Equivalence was assessed using the two one-sided tests procedure (TOST; Lakens, 2017), with equivalence bounds set at ±0.5 of the pooled standard deviation for each version pair. Item-level back-translation fidelity was assessed with paired *t*-tests, and Cohen's *d* comparing item means between T1 and T3.

Structural validity was examined on the Time 2 Pidgin responses. Sampling adequacy was checked with the Kaiser–Meyer–Olkin measure and Bartlett's test of sphericity before extraction was attempted. Items were then submitted to iterated principal axis factoring, which recovers common rather than total variance and is the appropriate choice where the question is whether a latent construct underlies the responses. Squared multiple correlations were used as initial communality estimates and iterated to convergence. Retention was decided by parallel analysis (Horn, 1965) rather than the eigenvalue-greater-than-one rule, which is known to overfactor; 1,000 random datasets matching the observed matrix in order were generated, and the observed eigenvalues were compared against the 95th percentile of the simulated distribution. Loadings of .40 and above were treated as salient, following the usual convention for exploratory work with modest samples (Stevens, 2009). Average variance extracted and composite reliability were then computed from the standardised loadings using the formulae of Fornell and Larcker (1981), against benchmarks of .50 and .70 respectively. A sample of 65 is small by the older participant-to-item rules of thumb, although MacCallum et al. (1999) demonstrated that these requirements fall sharply where communalities are high and few factors are expected, both of which hold here; the KMO value reported below offers a direct empirical check on that judgement. The analysis is nonetheless presented as preliminary evidence of structural validity rather than as a confirmation of dimensionality.

Ethical Considerations

The study was conducted in accordance with the ethical principles of the Declaration of Helsinki. Participants provided written informed consent before each wave, were informed of their right to withdraw at any point without penalty, and received no monetary inducement. Data were anonymised before analysis.

4. RESULTS

Descriptive Statistics

Item means and standard deviations for all 10 ESIS items across the three administration waves are reported in Table 1. Means ranged from 3.26 (item 8, T1) to 4.52 (items 6 and 7, T1). Several items showed negative skew at T1, consistent with ceiling effects that often arise when positively keyed well-being items are administered to a student population. The pattern was somewhat less pronounced at T2 and T3, a common feature of repeated administration. Total scale means were similar across versions: T1 M = 40.83 (SD = 7.60), T2 M = 38.51 (SD = 11.52), T3 M = 40.26 (SD = 9.31). The widening of dispersion at T2 is worth noting and is picked up later.

Table 1. *Item-Level Descriptive Statistics for the 10 ESIS Items Across the Three Versions*

	T1: Original English			T2: Pidgin English			T3: Back-Translated English		
Item	***M***	***SD***	***Skew***	***M***	***SD***	***Skew***	***M***	***SD***	***Skew***
SIS1	4.20	1.35	-1.52	3.46	1.72	-0.47	4.00	1.45	-1.18
SIS2	3.80	1.25	-1.13	3.57	1.47	-0.74	3.88	1.27	-0.80
SIS3	3.45	1.47	-0.55	3.85	1.36	-1.10	4.03	1.27	-1.23
SIS4	4.28	1.19	-1.76	3.86	1.32	-0.91	3.88	1.27	-0.90
SIS5	4.32	1.08	-2.00	4.00	1.41	-1.23	3.82	1.32	-1.12
SIS6	4.52	1.20	-2.46	4.12	1.62	-1.40	4.37	1.39	-1.86
SIS7	4.52	1.12	-2.27	4.15	1.54	-1.48	4.35	1.27	-1.89
SIS8	3.26	1.50	-0.29	3.60	1.45	-0.77	3.77	1.25	-0.74
SIS9	4.11	1.24	-1.43	4.06	1.32	-1.29	4.11	1.24	-1.53
SIS10	4.37	0.96	-2.12	3.83	1.35	-1.09	4.06	1.17	-1.27
Scale total	M = 40.83; SD = 7.60			M = 38.51; SD = 11.52			M = 40.26; SD = 9.31		

Note. M = item mean (scale: 1–5); SD = standard deviation; Skew = skewness. Scale total M and SD refer to the 10-item sum score (range: 10–50). N = 65.

Internal Consistency and Item-Total Statistics

Cronbach's alpha was .81 at T1, .93 at T2, and .90 at T3 (Table 2). All three estimates meet or comfortably exceed the commonly cited threshold of .70 for research-grade instruments, and the T2 estimate slightly exceeds the Finnish original of .89. Corrected item-total correlations generally sat in the .40 to .93 range, with one exception worth flagging. Item 3 ("I belong to a group or community that is important for me") had a low corrected item-total correlation of .12 at T1, and removal of this item would have raised alpha to .84. The same item performed more conventionally at T2 (r = .42) and T3 (r = .47), which may suggest that the Pidgin phrasing clarified an aspect of the item that was read ambiguously in English by the sample. Item 8 showed a similar though less dramatic pattern, with r = .35 at T1 and r = .34 at T3, rising to r = .53 at T2.

Table 2. *Corrected Item-Total Correlations and Cronbach's α if Item Deleted Across the Three Versions*

	T1: Original English		T2: Pidgin English		T3: Back-Translated	
Item	r_{it}	α_{-i}	r_{it}	α_{-i}	r_{it}	α_{-i}
SIS1	0.45	0.80	0.62	0.93	0.63	0.89
SIS2	0.44	0.80	0.77	0.92	0.82	0.88
SIS3	0.12 †	0.84	0.42	0.94	0.47	0.90
SIS4	0.44	0.80	0.78	0.92	0.73	0.88
SIS5	0.73	0.77	0.80	0.92	0.62	0.89
SIS6	0.77	0.77	0.93	0.91	0.71	0.88
SIS7	0.67	0.78	0.90	0.92	0.80	0.88
SIS8	0.35	0.82	0.53	0.94	0.34	0.91
SIS9	0.65	0.79	0.79	0.92	0.63	0.89
SIS10	0.53	0.79	0.84	0.92	0.73	0.88
Cronbach's α	**.81**		**.93**		**.90**	

Note. r_{it} = corrected item–total correlation; α_{-i} = Cronbach's α if item deleted. † r_{it} < .30, indicating a weak item–scale relationship. N = 65.

Structural Validity of the Pidgin Version

The Time 2 Pidgin responses were suitable for factor analysis. The Kaiser–Meyer–Olkin measure of sampling adequacy was .83, which Kaiser (1974) classifies as meritorious, and no individual item returned a measure of sampling adequacy below .75. Bartlett's test of sphericity was significant, $\chi^2(45) = 626.13$, $p < .001$, confirming that the correlation matrix departed from an identity matrix. The determinant of the matrix was .00003, which clears the .00001 threshold conventionally taken to signal extreme multicollinearity, though not by a wide margin.

Parallel analysis retained a single factor. The first observed eigenvalue of 6.19 stood well above the 95th percentile of the corresponding random eigenvalues (1.10), whereas the second observed value of 0.68 fell below its random counterpart of 0.78. The Kaiser criterion would have retained two factors, the second unreduced eigenvalue being 1.05, but this is precisely the overfactoring that parallel analysis corrects. A forced two-factor solution returned a second factor with a sum of squared loadings of 0.65 and no item loading on it alone, which supports the one-factor decision.

The one-factor principal axis solution is presented in Table 3. It accounted for 60.96% of the total variance. All ten items loaded above the .40 criterion, with loadings from .44 on item 3 to .97 on item 6. Communalities ranged from .19 to .94. The root mean square residual was .07, within the .08 ceiling usually treated as acceptable, and the largest single residual correlation was .18.

Two further indices were derived from the loadings. The average variance extracted was .61, above the .50 benchmark of Fornell and Larcker (1981), indicating that the latent factor accounted for more variance in its indicators than measurement error did. Composite reliability was .94, comfortably above the .70 criterion. Because the model is unidimensional and congeneric, this figure is also McDonald's omega, and it may be read alongside the Cronbach's alpha of .93 reported earlier. H5 was therefore supported, offering preliminary evidence that the Pidgin items measure a single social inclusion construct rather than merely correlating with one another.

Table 3. *Sampling Adequacy, Factor Loadings and Communalities for the 10 Pidgin ESIS Items at Time 2*

Item	Abbreviated item content	MSA	Factor loading	Communality (h^2)
1	What I do every day is significant	.86	.65	.42
2	I get positive feedback on what I do	.91	.78	.61
3	I belong to an important group	.94	.44	.19
4	Other people need me	.81	.79	.63
5	I can influence the course of my life	.95	.84	.70
6	I feel that my life has purpose	.77	.97	.94
7	I can strive for things important to me	.75	.94	.88
8	I get help when I really need it	.75	.55	.30
9	I feel trusted	.77	.81	.66
10	I can influence my living environment	.94	.87	.76

Note. N = 65. MSA = item measure of sampling adequacy. Extraction: iterated principal axis factoring, one factor retained by parallel analysis; no rotation was required. Item content is abbreviated; full stems appear in the Measure section. KMO = .83; Bartlett's $\chi^2(45) = 626.13$, $p < .001$. Eigenvalue = 6.10, accounting for 60.96% of total variance. Average variance extracted = .61; composite reliability = .94; Cronbach's α = .93.

Cross-Version Pearson Correlations

Pearson correlations between total scores across the three versions are presented in Table 4. All three pairs produced statistically significant positive correlations, though none reached the .70 threshold often used to claim strong cross-version equivalence (Mokkink et al., 2010). T1 and T2 correlated at $r = .43$ (95% CI [.21, .61]), T1 and T3 at $r = .42$ (95% CI [.19, .60]), and T2 and T3 at $r = .45$ (95% CI [.24, .63]). The three coefficients are near-identical, which is itself an interesting pattern. It may suggest that the three versions share a common underlying construct, but that each contributes some version-specific variance, which is exactly what a cross-cultural adaptation literature would predict.

Table 4. *Cross-Version Pearson Correlations for Total ESIS Scores with 95% Confidence Intervals*

Version pair	*r*	95% CI	*p*	Meets r ≥ .70?	Interpretation
T1 ↔ T2	.43	[.21, .61]	.0003***	No	Weak–moderate
T1 ↔ T3	.42	[.19, .60]	.0006***	No	Weak–moderate
T2 ↔ T3	.45	[.24, .63]	.0001***	No	Moderate

*Note. T1 = Original English; T2 = Pidgin English; T3 = Back-Translated English. 95% CIs computed via Fisher's r-to-z transformation. r ≥ .70 is the recommended threshold for cross-version score equivalence (Mokkink et al., 2010). p-values are two-tailed. *** p < .001.*

Score Interchangeability: ICC and Repeated-Measures ANOVA

The single-rater, two-way random-effects ICC(2,1) for total scores across the three versions was .42 (95% CI [.26, .56]), which falls just under the benchmark for moderate agreement and well below the benchmark for good agreement (Koo & Li, 2016). The average-measures ICC(2,k) was .68, the more appropriate figure if one intends to use an average of three versions as a single estimate, though such an average would rarely be useful in practice.

The RM-ANOVA on total scores returned a non-significant main effect of version, $F(2, 128) = 1.77$, $p = .174$, $\eta^2_p = .011$. Condition means were 40.83 at T1, 38.51 at T2, and 40.26 at T3. The non-significant omnibus test contrasts with patterns reported in some other AI-assisted Pidgin adaptations of positive well-being measures in similar populations, where Pidgin versions produced systematically inflated total scores. Bonferroni-corrected follow-up tests (Table 5) confirmed that no pairwise difference reached significance: T1 versus T2, $t(64) = 1.75$, $p = .257$, d = 0.22; T1 versus T3, $t(64) = 0.50$, $p = 1.00$, d = 0.06; T2 versus T3, $t(64) = -1.28$, $p = .616$, d = −0.16. The Cohen's d values are small to negligible across the board.

Table 5. *Bonferroni-Corrected Post-Hoc Paired t-Tests for Total ESIS Score Differences Between Versions*

Comparison	*MD*	*SDdiff*	95% CI for MD	*t(64)*	*p (Bonf.)*	*d*
T1 vs T2	2.32	10.73	[−0.33, 4.98]	1.	.257	0.22
T1 vs T3	0.57	9.25	[−1.72, 2.86]	0.50	1.00	0.06
T2 vs T3	−1.75	11.05	[−4.49, 0.98]	−1.28	.616	−0.16

Note. MD = mean difference (first version − second version); SDdiff = standard deviation of pairwise difference scores; CI = confidence interval for MD; d = Cohen's d for paired samples. p-values are Bonferroni-corrected for three comparisons (α = .05/3 ≈ .017). Effect-size conventions: |d| ≥ 0.20 small, ≥ 0.50 medium, ≥ 0.80 large.

Equivalence Testing

A null difference is not, on its own, the same as statistical equivalence (Lakens, 2017). The two one-sided tests confirmed equivalence for all three pairs. With equivalence bounds set at ±0.5 pooled SD, T1 versus T3 achieved equivalence with p(TOST) = .001, T2 versus T3 at p(TOST) = .007, and T1 versus T2 at p(TOST) = .030. This is a substantively stronger result than a simple non-significant mean difference would allow. Table 6 summarises the equivalence tests.

Table 6. *Equivalence Testing (TOST) Results for Total ESIS Score Comparisons Across Version Pairs*

Comparison	*MD*	Equiv. bound	*p (TOST)*	Equiv. established?	Interpretation
T1 vs T2	2.32	±4.88	.030	Yes	Equivalent
T1 vs T3	0.57	±4.25	.001	Yes	Equivalent
T2 vs T3	−1.75	±5.24	.007	Yes	Equivalent

Note. MD = mean difference in total scores. Equivalence bound set at ±0.5 × pooled SD of each pair. Equivalence is inferred when p(TOST) < .05, indicating the observed difference falls within the bounds. p-values reflect the more conservative (larger) of the two one-sided tests.

Back-Translation Fidelity

Item-level paired *t*-tests comparing the original English (T1) and back-translated English (T3) item means are reported in Table 7. Six of the ten items showed no significant mean shift between the original and the back-translation, suggesting reasonable semantic fidelity on those items. Four items differed significantly

at the item level: item 3 (MD = −0.59, $t(64) = -3.17$, $p = .002$, d = −0.39), item 4 (MD = 0.40, $t(64) = 2.30$, $p = .025$, d = 0.29), item 5 (MD = 0.51, $t(64) = 2.56$, $p = .013$, d = 0.32), and item 8 (MD = −0.51, $t(64) = -2.48$, $p = .016$, d = −0.31). These tests are exploratory rather than confirmatory, and no correction for multiple comparisons was applied, though the effect sizes are all small. Taken together, these four items are plausible candidates for expert review in a subsequent iteration.

Table 7. *Item-Level Back-Translation Fidelity: Paired t-Tests Comparing Original English (T1) and Back-Translated English (T3) Item Means*

	T1: Original		T3: Back-Translated						
Item	***M***	***SD***	***M***	***SD***	***MD***	**95% CI**	***t(64)***	***p***	***d***
SIS1	4.20	1.35	4.00	1.45	0.20	[−0.24, 0.64]	0.91	.368	0.11
SIS2	3.80	1.25	3.88	1.27	-0.08	[−0.40, 0.25]	-0.47	.639	-0.06
SIS3	3.45	1.47	4.03	1.27	-0.59	[−0.95, −0.22]	-3.17	.002**	-0.39
SIS4	4.28	1.19	3.88	1.27	0.40	[0.05, 0.75]	2.30	.025*	0.29
SIS5	4.32	1.08	3.82	1.32	0.51	[0.11, 0.91]	2.56	.013*	0.32
SIS6	4.52	1.20	4.37	1.39	0.15	[−0.27, 0.58]	0.72	.472	0.09
SIS7	4.52	1.12	4.35	1.27	0.17	[−0.21, 0.55]	0.89	.378	0.11
SIS8	3.26	1.50	3.77	1.25	-0.51	[−0.92, −0.10]	-2.48	.016*	-0.31
SIS9	4.11	1.24	4.11	1.24	0.00	[−0.39, 0.39]	0.00	1.000	0.00
SIS10	4.37	0.96	4.06	1.17	0.31	[−0.08, 0.69]	1.59	.117	0.20

*Note. MD = mean difference (T1 − T3); CI = 95% confidence interval for MD; d = Cohen's d for paired samples. Items in bold showed statistically significant differences (p < .05), indicating potential semantic shift in the Pidgin English translation. These tests are exploratory; no correction for multiple comparisons was applied, and findings should be treated as indicative rather than confirmatory. p < .05. * p < .01.*

5. DISCUSSION

The study set out to evaluate an AI-assisted Pidgin version of the ESIS in a within-subject three-wave design with 65 Nigerian emerging adults, and the findings are more encouraging than we had anticipated going in. Internal consistency was excellent in the Pidgin version (α = .93) and acceptable to strong in both English administrations (α = .81 at T1 and .90 at T3). The Pidgin alpha exceeds the Finnish original of .89 reported by Leemann et al. (2022), which may reflect the relative homogeneity of the student sample, or the possibility that Pidgin reduced formal-register ambiguity present in the English items. Both mechanisms deserve investigation in a larger and more cognitively diverse sample.

The factor analysis adds a layer of evidence that the reliability coefficients on their own could not supply. A high alpha is consistent with a unidimensional scale but does not establish one, since alpha rises with the number of items and with any source of common variance. Parallel analysis on the Pidgin data retained one factor accounting for roughly 61% of the variance, every item loaded above .40, and the average variance extracted of .61 indicates that the factor carried more of the item variance than error did. This matches the single-factor structure Leemann et al. (2022) reported for the Finnish original, which is the more useful comparison. The Pidgin version is not merely internally consistent; it appears to be internally consistent about the same thing. The fifth hypothesis was therefore supported.

Two items qualify that reading. Item 3, on belonging to a group or community that matters to the respondent, was the weakest indicator at Time 2, with a loading of .44 and a communality of .19, meaning that better than four-fifths of its variance sat outside the common factor. This is the same item that returned a corrected item-total correlation of .12 at Time 1 and one of the four significant back-translation shifts. Three independent analyses now converge on item 3, which makes the case for targeted rewording rather stronger than a single anomalous statistic would. Item 8, on receiving help when it is needed, was the second weakest on each of these indices.

A different issue appears at the other end of the distribution. Items 6 and 7, on life having purpose and on being able to strive for things that matter, correlated at .96 in the Pidgin version, against .78 in the original English and .87 in the back-translation, and they carried the two highest loadings. One reading is that the Pidgin renderings of these two stems are not fully discriminable to respondents, which would inflate the Time 2 alpha and would go some way towards explaining why that estimate exceeded the Finnish original. Near-redundant items add length without adding information, so this is worth attention in any revision.

Cross-version correlations sat in the .42 to .45 range, below the .70 benchmark recommended for strong score equivalence (Mokkink et al., 2010). At first glance these coefficients read as modest. Read in context, however, they are slightly more interesting. The near-identical magnitude of all three correlations is not what one would expect if the Pidgin version were materially distorting the construct. If Pidgin were unusually noisy or systematically biased, we should see a weaker T1–T2 correlation than T1–T3. Instead the three pairs cluster tightly. It may therefore be the case that each version contributes a small, version-specific layer of variance, perhaps attributable to translation, mode of processing, or occasion-related factors, but that the central construct is preserved across all three.

The absence of a significant main effect of version on total scores is the clearest contrast with some related adaptation work. Several AI-assisted Pidgin adaptations of eudaimonic well-being measures have reported systematic score inflation in the Pidgin version with medium-to-large effect sizes, which can seriously compromise cross-version interchangeability. The present ESIS data showed no such pattern, and all pairwise Cohen's d values were small to negligible. What is more, the two one-sided tests procedure confirmed statistical equivalence for each of the three pairs within the ±0.5 pooled SD boundary, with the tightest equivalence observed for T1 versus T3 ($p = .001$). The ESIS appears to have weathered AI-assisted translation without the inflation that has sometimes surfaced for more culturally loaded eudaimonic constructs.

One plausible reading is that the ESIS's relatively short item stems and its concrete framing of meaningful daily activity, positive feedback, group belonging, and helping and being helped lend themselves to Pidgin rendering without forcing the translator to improvise vocabulary for abstract psychological states. Instruments with more formally phrased eudaimonic content may place greater strain on an LLM that is otherwise competent at everyday Pidgin usage. This is speculative, though it fits the comparative pattern with earlier adaptation work in the same population.

The moderate ICC(2,1) of .42 remains a source of caution. Koo and Li's (2016) benchmarks place this value near the lower end of moderate agreement, meaning that although the three versions produce non-different means that are formally equivalent, they still disagree at the individual level more than one would want for confident interchange. A researcher comparing, for example, the Pidgin scores of one cohort with the English scores of another would be likely to introduce measurement noise that could obscure real group differences. Score equivalence at the group level does not automatically license person-level score interchangeability, a point that is easy to miss in the adaptation literature.

Four items showed small but statistically detectable mean shifts between the original and the back-translation: items 3 (belonging to a group or community), 4 (other people need me), 5 (influence on the course of one's life), and 8 (getting help when needed). These do not cluster into a single theoretical grouping, which is slightly reassuring in that the drift appears item-specific rather than systemic. Item 3 in particular had a low corrected item-total correlation of .12 at T1, which rose to .42 at T2 and .47 at T3. A tentative reading is that the English phrasing "I belong to a group or community that is important for me" was read with unexpected ambiguity by respondents in this sample, perhaps because the word "community" carries a range of referents in Nigerian English, encompassing kinship, church, age grade, and online group. The Pidgin version may, in effect, have narrowed this ambiguity. If so, the problem item is not the Pidgin item but the original English item as received by this sample. Cognitive interviewing would be the clearest way to adjudicate.

The present findings sit within a wider debate on what AI-assisted translation can and cannot do for low-resource languages. The results here are broadly compatible with Grobelny et al. (2025) on Polish LLM adaptation and with the non-inferiority findings of Sørensen et al. (2025). They are less compatible with the warning from AlKhamissi et al. (2024) on cultural misalignment, though that literature was concerned with open-ended generation rather than focused translation. Our more modest claim is that an AI-built Pidgin ESIS appears to behave like a plausible draft version and warrants a properly powered validation study before any broader claim is advanced.

Limitations

Three caveats are worth registering. First, the sample is small (N = 65), drawn from a single institution, and consists entirely of fourth-year psychology students. The exploratory factor analysis reported here is

defensible at this size given the high communalities and the single-factor solution recovered (MacCallum et al., 1999), and the KMO value of .83 supports that judgement empirically, but it remains an exploratory result in one sample. Confirmatory factor analysis, tests of measurement invariance across the three versions, and full cross-cultural structural comparison would each require an independent and substantially larger sample, and a mature adaptation programme would need to include them. Second, the within-subjects design cannot fully separate version effects from carryover. A two-week interval seems unlikely to preserve verbatim memory in any strong form, but we cannot rule out that earlier responses shaped later ones in subtle ways. A parallel-samples design, with different respondents receiving different versions, would resolve this at the cost of losing the statistical power that within-subjects analyses provide. Third, the translation pipeline relied exclusively on AI applications, without independent bilingual human review of each item. The four items that showed significant T1–T3 shifts are consistent with the possibility that this process produced localised translation artefacts rather than genuine construct drift. A subsequent validation wave with expert committee review and cognitive debriefing would address this more convincingly.

6. CONCLUSION

Against its modest aspirations, the Nigerian Pidgin ESIS performed well. Internal consistency was strong across all three administrations, the Pidgin items resolved into the single factor reported for the original instrument, every pair of versions reached formal statistical equivalence, and only four of the ten items showed detectable item-level drift. The moderate intraclass correlation flags person-level score interchangeability as still unresolved, and no serious claim of full validation can be made from a single pilot with 65 respondents. Even so, the evidence presented here is sufficient to justify moving to a properly powered follow-up study. Nigerian Pidgin, which is the everyday language of a large share of Nigeria's emerging adult population, is an appropriate vehicle for a well-built social inclusion instrument, and the ESIS appears to adapt to this language without the score distortions that have attended some earlier AI-assisted adaptations of more eudaimonic instruments.

7. RECOMMENDATIONS

Four lines of follow-up suggest themselves on the basis of the present findings. First, items 3, 4, 5, and 8 should be examined in an expert committee that includes bilingual psychologists and linguists who are native Pidgin users. Minor rewording at these four sites, rather than a wholesale rework, is likely sufficient. Items 6 and 7 should also be placed before the same committee. Their near-perfect correlation in the Pidgin version suggests that the two stems may have collapsed into a single meaning in translation, and separating them would improve the information the scale carries per item. Second, cognitive interviewing with a purposive sample of 10 to 12 Pidgin speakers across age and educational strata would help clarify whether the shifts observed in item 3 reflect Pidgin phrasing or a latent ambiguity in the English source item. Third, a confirmatory study with an independent-groups design and a minimum of 200 participants per language condition would support confirmatory factor analysis, tests of configural, metric, and scalar measurement invariance, and test–retest reliability over longer intervals. Fourth, that study should include measures of related constructs such as perceived social support, loneliness, and quality of life to establish convergent and discriminant validity, along with demographic variables detailed enough to permit differential item functioning analysis by age, gender, and educational background. If these steps are taken, the ESIS in Nigerian Pidgin could find a proper place in the small but growing set of psychometric instruments available in the language most widely spoken in everyday Nigerian life.

REFERENCES

Ajagbe, S., Fonkeu, B., & Edegbe, U. (2022). West African Pidgin as a tool for socio-economic development. *Language Problems and Language Planning, 46*(1), 1–25. https://doi.org/10.1075/lplp.21030.aja

AlKhamissi, B., ElNokrashy, M., AlKhamissi, M., & Diab, M. (2024). Investigating cultural alignment of large language models. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics* (Vol. 1, pp. 12404–12422). Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.acl-long.671

Arnett, J. J. (2000). Emerging adulthood: A theory of development from the late teens through the twenties. *American Psychologist, 55*(5), 469–480. https://doi.org/10.1037/0003-066X.55.5.469

Arnett, J. J. (2015). *Emerging adulthood: The winding road from the late teens through the twenties* (2nd ed.). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199795574.013.9

Bartolomei, L., Blundo-Canto, G., & De Muro, P. (2024). How is the Capability Approach Applied to Assess Well-being Impacts? A Systematic Review. *Journal of Human Development and Capabilities*, *25*(3), 367–399. https://doi.org/10.1080/19452829.2024.2369502

Beaton, D. E., Bombardier, C., Guillemin, F., & Ferraz, M. B. (2000). Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine, 25*(24), 3186–3191. https://doi.org/10.1097/00007632-200012150-00014

Chukwu, J., Ukwaja, K. N., Murphy-Okpala, N., Ekeke, N., Eze, C., Iyama, F., Meka, A., Njoku, M., Ezeakile, O., Dahiru, T., Abdullahi, S., Anyaike, C., & Nwafor, C. (2023). Translation, cross-cultural adaptation and validation of the Nigerian Pidgin English version of the Rosenberg Self-Esteem Scale and the Leprosy-Adapted Internalized Stigma of Mental Illness Scale among persons affected by leprosy in southern Nigeria. *medRxiv*. https://doi.org/10.1101/2023.07.01.23292117

Chukwuorji, J. B. C., & Osondu, O. (2023). Translation of well-being assessment instruments in African contexts: A mapping review and future directions. *Journal of Psychological Research, 5*(4), 15–35. https://doi.org/10.30564/jpr.v5i4.5907

Counted, V., Cowden, R. G., & Lomas, T. (2024). Multidimensional flourishing in Africa: An intracontinental analysis of 38 well-being indicators in 40 countries. *Journal of Happiness Studies, 25*, Article 51. https://doi.org/10.1007/s10902-024-00764-5

Deuber, D. (2005). *Nigerian Pidgin in Lagos: Language contact, variation and change in an African urban setting*. Battlebridge Publications.

Eberhard, D. M., Simons, G. F., & Fennig, C. D. (Eds.). (2024). *Ethnologue: Languages of the World* (27th ed.). SIL International. https://www.ethnologue.com

Elugbe, B. O., & Omamor, A. P. (1991). *Nigerian Pidgin: Background and prospects*. Heinemann Educational Books.

Fadele, K. P., Igwe, S. C., Udokang, E., Niji-Olawepo, T., Ogaya, J. B., & Lucero-Prisno, D. E., III. (2024). Mental health challenges in Nigeria: Bridging the gap between demand and resources. *Cambridge Prisms: Global Mental Health, 11*, Article e29. https://doi.org/10.1017/gmh.2024.19

Faraclas, N. (1996). *Nigerian Pidgin*. Routledge.

Fatoye, F., Akinfala, A. E., Gebrye, T., Fatoye, C., Ojelade, T. F., Oyewole, O. O., & Mbada, C. E. (2022). Translation, cross-cultural adaptation, and psychometric testing of Yoruba version of the EQ-5D questionnaire in patients with musculoskeletal disorders. *Frontiers in Public Health, 10*, Article 902680. https://doi.org/10.3389/fpubh.2022.902680

Google. (2026). *Gemini AI* [Large language model]. https://gemini.google.com

Grobelny, J., Szymański, K., & Strozyk, Z. (2025). Act as an expert in psychometry: The evaluation of large language models utility in psychological tests cross-cultural adaptations. *Acta Psychologica, 261*, 105813. https://doi.org/10.1016/j.actpsy.2025.105813

Guillemin, F., Bombardier, C., & Beaton, D. (1993). Cross-cultural adaptation of health-related quality of life measures: Literature review and proposed guidelines. *Journal of Clinical Epidemiology, 46*(12), 1417–1432. https://doi.org/10.1016/0895-4356(93)90142-N

Hamza, A. M., Nabilla, B. A. M. A.-S., & Low, Y. S. (2013). Measuring perceived social support in stroke survivors: Linguistic validation of the Multidimensional Scale of Perceived Social Support (MSPSS) in Hausa (Nigerial) Language. *South African Journal of Occupational Therapy*, *42*(3), 26-29. https://journals.assaf.org.za/index.php/sajot/article/view/19719

Igwesi-Chidobe, C. N., Kitchen, S., Sorinola, I. O., & Godfrey, E. L. (2021). Adaptation and validation of the Nigerian Igbo Multidimensional Scale of Perceived Social Support in patients with chronic low back pain. *Measurement Instruments for the Social Sciences, 3*, Article 8. https://doi.org/10.1186/s42409-021-00026-7

Kaiser, B. N., Ticao, C., Anoje, C., Boglosa, J., Gafaar, T., Minto, J., & Kohrt, B. A. (2022). Challenges in simultaneous validation of mental health screening tools in multiple languages: Adolescent assessments in Hausa and Pidgin in Nigeria. *SSM – Mental Health, 2*, Article 100168. https://doi.org/10.1016/j.ssmmh.2022.100168

Khumalo, I. P., Temane, Q. M., & Wissing, M. P. (2010). Development and initial validation of a General Psychological Well-Being Scale (GPWS) in an African context. *Journal of Psychology in Africa, 20*(1), 13–22. https://doi.org/10.1080/14330237.2010.10820338

Khumalo, I. P., Temane, Q. M., & Wissing, M. P. (2013). Further validation of the General Psychological Well-Being Scale among a Setswana-speaking group. In M. P. Wissing (Ed.), *Well-being research in South Africa* (pp. 199–224). Springer. https://doi.org/10.1007/978-94-007-6368-5_10

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine, 15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Lakens, D. (2017). Equivalence tests: A practical primer for t tests, correlations, and meta-analyses. *Social Psychological and Personality Science, 8*(4), 355–362. https://doi.org/10.1177/1948550617697177

Leemann, L., Martelin, T., Koskinen, S., Harkko, J., & Isola, A.-M. (2022). Development and psychometric evaluation of the Experiences of Social Inclusion Scale. *Journal of Human Development and Capabilities, 23*(3), 400–424. https://doi.org/10.1080/19452829.2021.1985440

Li, H., Zhang, H., Li, M., Wang, Y., Wen, L., Zhang, Y., & Huang, B. (2026). Toward robust multilingual adaptation of LLMs for low-resource languages. *arXiv*. https://doi.org/10.48550/arXiv.2510.14466

Lu, S. C., Xu, C., Kaur, M., Agarwal, A., Dedic, A., Schmit, E., Xu, L., Carlino, G., Tsangaris, E., Gibbons, C., & Pusic, A. L. (2025). Can machine translation match human expertise? Quantifying the performance of large language models in the translation of patient-reported outcome measures (PROMs). *Journal of Patient-Reported Outcomes, 9*, Article 94. https://doi.org/10.1186/s41687-025-00926-w

Mbada, C. E., Awokoya, A. S., Oyewole, O. O., Idowu, O. A., Akinsulore, A., Fatoye, C., Oke, K. I., & Fatoye, F. (2021). Translation, cross-cultural adaptation and psychometric evaluation of yoruba version of the short-form 12 health survey. *Annali di igiene : medicina preventiva e di comunita*, *33*(3), 254–267. https://doi.org/10.7416/ai.2021.2431

Mgbeojedo, U. G., Akosile, C. O., Ezugwu, J. C., Okoye, E. C., John, J. N., Ani, K. U., & Okezue, O. C. (2022). Cross-cultural adaptation and validation of the 15-item Geriatric Depression Scale (GDS-15) into Igbo language: A validation study. *Health and Quality of Life Outcomes, 20*(1), Article 19. https://doi.org/10.1186/s12955-022-01928-8

Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., Bouter, L. M., & de Vet, H. C. W. (2010). The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *Journal of Clinical Epidemiology, 63*(7), 737–745. https://doi.org/10.1016/j.jclinepi.2010.02.006

Murphy-Okpala, N., Dahiru, T., Ukwaja, K. N., Van't Noordende, A., Eze, C., Nwafor, C., Iyama, F., Van Brakel, W., Schoenmakers, A., Meka, A., Ekeke, N., Abdullahi, S., Njoku, M., Ezeakile, O., Bassey, J. O., Anyaike, C., Sesere, O., & Chukwu, J. (2025). Cross-cultural validation of the Pidgin-English version of the EMIC-CSS and SDS among adults in Southern Nigeria. *Leprosy Review, 96*(3), Article e2025095. https://doi.org/10.47276/lr.96.3.2025095

Nussbaum, M. C. (2011). *Creating capabilities: The human development approach*. The Belknap Press of Harvard University Press. https://doi.org/10.2307/j.ctt2jbt31

Obidoa, C. A., Dodor, B. A., Tackie-Ofosu, V., Obidoa, M. A., Kalisch, H. R., & Nelson, L. J. (2019). Perspectives on markers of adulthood among emerging adults in Ghana and Nigeria. *Emerging Adulthood, 7*(4), 270–278. https://doi.org/10.1177/2167696818778906

Odole, A. C., & Akinpelu, A. O. (2009). Translation and alternate forms reliability of the Visual Analogue Scale in the three major Nigerian languages. *Internet Journal of Allied Health Sciences and Practice, 7*(3), Article 13. https://doi.org/10.46743/1540-580X/2009.1257

Oversabi. (2026). *Oversabi AI* [AI language application]. https://oversabi.app/chat

Rotzinger, J. S., Arnett Jensen, L., & Thalmayer, A. G. (2025). Ubuntu in Namibia and Kenya: How emerging adults live an essential African value today. *Journal of Cross-Cultural Psychology, 56*(2–3), 165–188. https://doi.org/10.1177/00220221241309863

Sen, A. (1999). *Development as freedom*. Oxford University Press.

Sørensen, C. S. D., Damsgaard, T. M., Andresen, K., Rosenberg, J., & Baker, J. J. (2025). Comparing ChatGPT-4 and human translation of an outcome questionnaire: A randomized, double-blinded non-inferiority study. *Cureus, 17*(4), Article e82525. https://doi.org/10.7759/cureus.82525

Thalmayer, A. G., Asatsa, S., Shino, E. N., Naudé, L., Laher, S., Hassem, T., Florence, M., Adonis, T.-A., Uugwanga, S. N., Rotzinger, J. S., Hofmann, D., Makunda, J., Botha, C., Murangi, A., & Shirima, C. M. (2024). Lifespan research in Kenya, Namibia, and South Africa: Cohort profile of the Africa Long Life Study. *Methods in Psychology, 11*, Article 100153. https://doi.org/10.1177/27000710241264492

Wang, J., Liang, Y., Sun, Z., Cao, Y., Xu, J., & Meng, F. (2024). Cross-lingual knowledge editing in large language models. In L.-W. Ku, A. Martins, & V. Srikumar (Eds.), *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 11676–11686). Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.acl-long.627

Wild, D., Grove, A., Martin, M., Eremenco, S., McElroy, S., Verjee-Lorenz, A., & Erikson, P. (2005). Principles of good practice for the translation and cultural adaptation process for patient-reported outcomes (PRO) measures: Report of the ISPOR Task Force for Translation and Cultural Adaptation. *Value in Health, 8*(2), 94–104. https://doi.org/10.1111/j.1524-4733.2005.04054.x

Yakpo, K. (2024). West African Pidgin: World language against the grain. *Africa Spectrum, 59*(2), 180–203. https://doi.org/10.1177/00020397241263364

ACKNOWLEDGEMENT

We acknowledge and appreciate the Tertiary Education Trust Fund (TETFund) for sponsoring this research project under the 2024 National Research Fund (NRF).