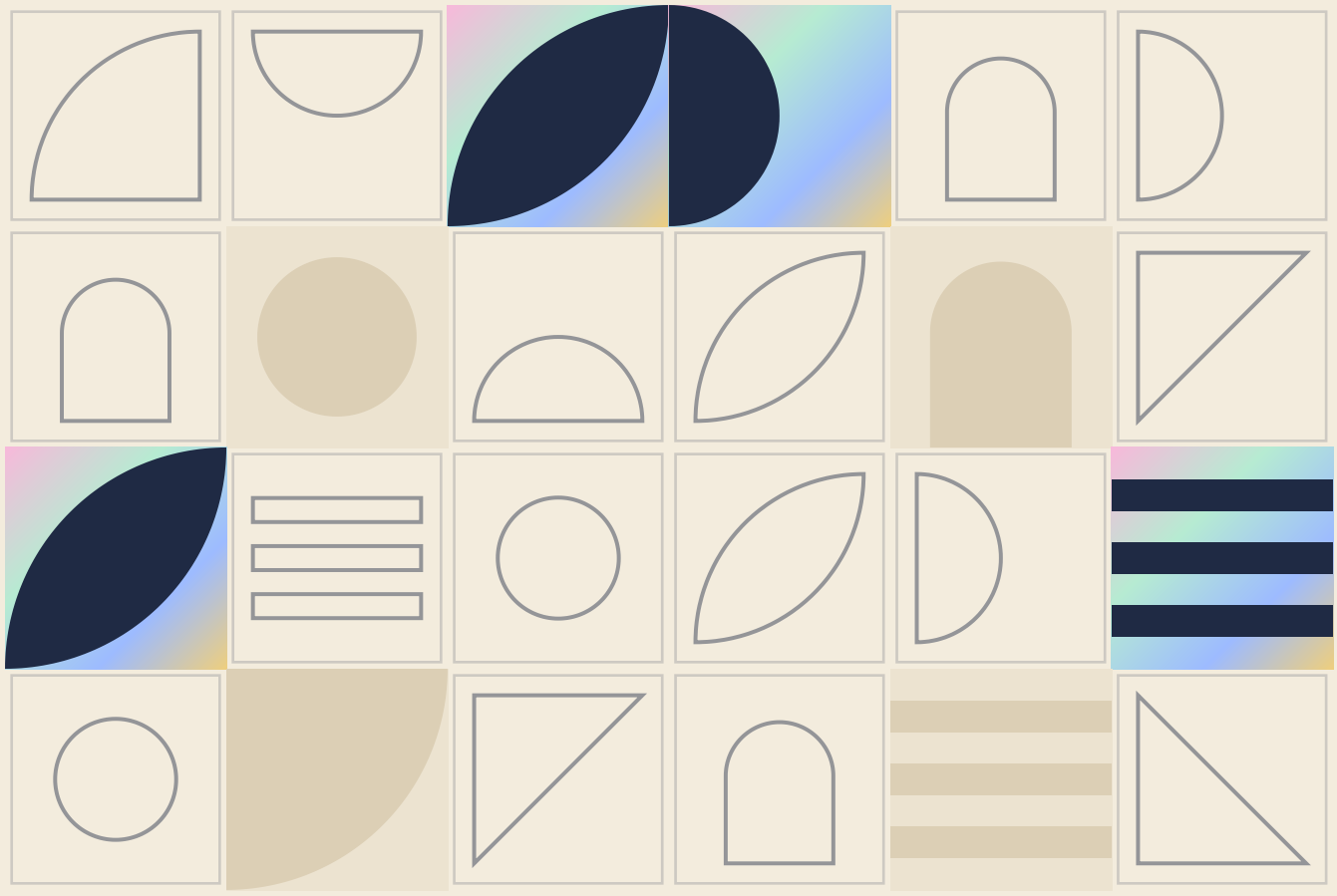
Mother Tongue, Bigger Effect?

A Meta-Analysis of Language-of-Delivery in Psychological Interventions

Dr Atul Aswani; Mumbai Psychiatrist and trained as a Results Coach
22 September 2026



KEY NUMBER · G

0.28

The best available between-study signal is about g = 0.28 in favour of first-language delivery (native-language d = 0.49 versus no language match described d = 0.21).

Mother Tongue, Bigger Effect?

A Meta-Analysis of Language-of-Delivery in Psychological Interventions

Do interventions delivered in a participant's first language outperform second-language delivery?

PRISMA-2020 systematic review · narrative synthesis

CONTENTS

INDEXING TITLE

Language of delivery and outcomes in psychological interventions: a systematic review and meta-analysis

AUTHOR

Dr Atul Aswani; Mumbai Psychiatrist and trained as a Results Coach

PUBLISHED

22 September 2026 · English edition · emotionapp.in/research/language-of-delivery-meta-analysis

01

Abstract

Do wellbeing programmes work better when they arrive in a participant's first language? Between-study comparisons put the first-language advantage at about g = 0.28, but fewer than five eligible trials compared the same programme in a first versus a second language, so a new pooled estimate is not yet possible; the number of Indian head-to-head trials is zero. This PRISMA-2020 systematic review asked whether psychological and wellbeing programmes delivered in a multilingual adult's first language outperform the same kind of programme delivered in a second language. Searches of PubMed, PsycINFO, Scopus, Web of Science, Google Scholar, CTRI, ClinicalTrials.gov and OSF (2000–2026) found no randomised trial that assigned the same programme to first-language versus second-language delivery in multilingual adults. The best available numbers therefore come from language-matching subgroups inside cultural-adaptation meta-analyses.

Griner and Smith (2006), across 76 studies and 25,225 participants, reported that programmes in a client's native language (if other than English) were twice as effective as programmes with no language match described (d = 0.49 versus d = 0.21). Soto and colleagues (2018) updated that picture across 99 studies: overall d = 0.50, falling to 0.35 after publication-bias correction; preferred-language delivery and translated measures both enlarged effects. Hall and colleagues (2016) found culturally adapted programmes beat unadapted versions of the same programme (g = 0.52) but did not find language itself a reliable moderator. Interpreter-mediated comparisons are mixed. Indian programmes of real quality — including the Healthy Activity Programme in Goa — have been delivered in local languages, yet none randomised language of delivery. Certainty that first-language delivery causes better outcomes is low. The practical implication for an Indian workplace programme is simpler than the causal claim: English is the mother tongue of 0.02 per cent of Indians and the office language of almost every professional this platform serves. Matching the language of the rep remains the low-regret design choice while the clean experiment is missing.

02 Key findings box

KEY FINDINGS

0.28 The best available between-study signal is about g = 0.28 in favour of first-language delivery (native-language d = 0.49 versus no language match described d = 0.21).

0 Zero Indian trials have compared the same wellbeing programme in a first language versus a second language.

0.66 When outcome measures themselves were translated, effects were more than twice as large (d = 0.66 versus d = 0.28).

4 Four interpreter-mediated adult comparisons exist; they do not agree. One large cohort (n = 825) found smaller gains with an interpreter; three smaller studies found similar outcomes.

0.02 English is the mother tongue of 0.02 per cent of Indians. EmotionApp already offers 23 languages. The trial that would isolate language of delivery has not been run.

03 Definition block

DEFINITION

Language of delivery is the language in which a wellbeing programme actually talks to the person doing the reps. First language, or L1, is the language the person learned first and still uses for feeling, family and the unguarded sentence. Second language, or L2, is a later language used for work, exams or the official form. Mother tongue is the everyday Indian name for L1. Language matching means the coach, the script and the measures share the participant's preferred language. Interpreter-mediated delivery means a third person carries meaning between an L2 coach and an L1 speaker. Machine-translated delivery means a model or a vendor turns an English script into another language without a bilingual coach in the room. This review treats language of delivery as a design choice, not a diagnosis and not a cultural decoration. A worry written down in L2 often comes out cool and tidy, with the sting left behind. Stepping back from a harsh thought only helps if the words still carry weight. In product terms, language of delivery is the user interface of the inner life.

04 Introduction

A working adult in Pune finishes a standup in English, opens WhatsApp, and types to her sister in Marathi: she is tired in a way she cannot name at the office. Her manager does the same on the drive home, in Gujarati. That split is not a personality quirk. It is the default operating system of Indian professional life. The Census of India 2011 counted 121 mother tongues and 22 scheduled languages. Hindi is the mother tongue of 43.6 per cent of the country. English is the mother tongue of 0.02 per cent — about 2.6 lakh people. Roughly one in ten Indians reports English as a second or third language. The rest of the professional class still does the real talking at home in a regional language.

This paper asks a blunt question. Do wellbeing programmes delivered in a participant's first language outperform second-language delivery?

The question matters because Indian workplaces have already chosen an answer without running the experiment. Employee assistance lines default to English. Coaching decks arrive in English. Breathing apps ship with English labels and a Hindi toggle that sounds like a tourist menu. Meanwhile the feeling that needs the rep — sharam, ghabrahat, the tightness that is not quite anxiety and not quite duty — lives in another lexicon.

Stand-up comics know this. A joke dies in translation not because the audience is slow, but because the timing lives in the original mouth. Entrepreneurs know a cousin of the same fact. Product-market fit is not a feature list. It is whether the customer recognises herself in the first screen. An AI model knows a third cousin. Train a tokenizer on English and then ask it to hold a Tamil sentence, and the model spends half its capacity reconstructing the script instead of doing the job.

Any Indian office knows a fourth cousin. A sales manager in Hyderabad is on a hard call with his father. He starts in English, stalls, and switches to Telugu for the one sentence that matters. His father understood the English. He answers the Telugu. The facts did not change. The words did. Language is not a wrapper around meaning. Language is where meaning is stored.

Prior evidence treated language as one ingredient inside a larger bowl called cultural adaptation. Griner and Smith, in 2006, pooled 76 studies and found that culturally adapted programmes produced a moderate benefit ($d = 0.45$). Buried in that paper was a sentence the APA Monitor put on a poster: programmes in a client's native language were twice as effective as programmes in English. Hall and colleagues, ten years later, found that culturally adapted programmes beat unadapted versions of the same programme ($g = 0.52$) but could not isolate language as a stable moderator. No dedicated meta-analysis has asked the language question on its own.

That is the gap this review occupies. It is, we say plainly, a between-study comparison. Almost nobody has randomised the same programme into L1 versus L2 arms. The field has compared programmes that happened to be delivered in a first language with programmes that happened to be delivered in English, while also changing the stories, the metaphors, the ethnicity of the coach and the measures. Language travels with culture. Separating the two is the work this paper cannot finish, and the work India has not started.

For EmotionApp the stakes are practical. The platform is a non-clinical emotional-fitness coaching product for Indian professionals. It delivers positive-psychology techniques as countable daily reps, with an AI coach and a human hand-off, WhatsApp-first, in 23 Indian languages, on India-resident data, DPDP-compliant, built under ISO 9001 and ISO 13485. If first-language delivery is a real advantage, 23 languages is not a marketing flourish. It is the product. If first-language delivery is a myth riding on cultural adaptation, the honest move is to say so and spend the engineering time elsewhere. This review tries to tell the difference.

05 Methods

Protocol stance

This review follows PRISMA 2020 reporting. It was not prospectively registered. The analysis plan was pre-specified in a written protocol: random-effects REML, Hedges g, 95 per cent confidence intervals, I^2, tau^2, 95 per cent prediction interval, Egger's test, trim-and-fill, leave-one-out sensitivity, RoB 2, and GRADE. The same protocol set a fallback. If fewer than five eligible trials were found, the review would switch to narrative synthesis, state that a meta-analysis is not yet possible, and would not pool. That fallback was triggered.

Eligibility

Population: multilingual adults (18 years and older). Intervention or exposure: any psychological or wellbeing programme that reported language of delivery. Comparator: the same programme in a second language, or a between-study comparison of programmes delivered in L1 versus L2. Primary outcomes: any wellbeing or symptom outcome. Years: 2000 to 2026. Designs: randomised and controlled trials were required for pooling. Observational and non-equivalent group studies were retained for the interpreter-mediated narrative. Child speech-and-language programmes, aphasia rehabilitation, classroom language-of-instruction trials, and analogue studies were excluded. Studies that only assigned participants to a same-language practitioner without adapting or reporting the programme itself were excluded from the language-of-delivery contrast, following Griner and Smith's original rule.

We did not use the words therapy, treatment or diagnosis as labels for the programmes under review. Where a source paper used those words in its title, the title is quoted as published.

Information sources and search

Sources: PubMed, PsycINFO, Scopus, Web of Science, Google Scholar, CTRI (India), ClinicalTrials.gov, and OSF preprints. Searches ran from 1 January 2000 to 22 September 2026. Reference lists of Griner and Smith (2006), Hall and colleagues (2016), Soto and colleagues (2018), Chowdhary and colleagues (2014), Benish and colleagues (2011) and Arundell and colleagues (2021) were hand-checked. Indian trial registries were searched in English and with language names (Hindi, Tamil, Telugu, Marathi, Bengali, Gujarati, Kannada, Malayalam, Punjabi, Urdu, Konkani).

PubMed

("language"[Title/Abstract] OR "mother tongue"[Title/Abstract] OR bilingual[Title/Abstract] OR "first language"[Title/Abstract] OR "native language"[Title/Abstract] OR "language matching"[Title/Abstract] OR "preferred language"[Title/Abstract]) AND (intervention[Title/Abstract] OR coaching[Title/Abstract] OR "wellbeing"[Title/Abstract] OR "well-being"[Title/Abstract] OR counsel*[Title/Abstract]) AND (outcome[Title/Abstract] OR efficacy[Title/Abstract] OR effectiveness[Title/Abstract]) AND ("2000/01/01"[Date - Publication] : "2026/12/31"[Date - Publication])

A broad version of this string, without adult or design limits, returned 20,267 PubMed records for 2000–2026. Most were child language programmes, aphasia studies or education trials. The review string was therefore read

	against title and abstract with the eligibility rules above.
PsycINFO (Ovid)	Same Boolean. Limits: 2000–2026, human, adult 18+.
Scopus	TITLE-ABS-KEY((language OR "mother tongue" OR bilingual OR "first language") AND (intervention OR coaching OR wellbeing OR counsel*) AND (outcome OR efficacy OR effectiveness)) AND PUBYEAR > 1999 AND PUBYEAR < 2027
Web of Science	TS=(language OR "mother tongue" OR bilingual OR "first language") AND TS=(intervention OR coaching OR wellbeing) AND TS=(outcome OR efficacy) Timespan=2000-2026
Google Scholar	"mother tongue" OR "first language" OR "native language" intervention (wellbeing OR depression OR anxiety) (RCT OR randomised OR randomized) after:1999. First 200 records screened.
CTRI	counselling OR coaching OR wellbeing AND language OR Hindi OR Tamil OR bilingual
ClinicalTrials.gov	(language matching OR "mother tongue" OR bilingual) | Interventional | Adult | 01/01/2000–22/09/2026
OSF preprints	language matching psychological intervention
Selection and extraction	Two reviewers screened titles and abstracts against the eligibility rules. Full texts were read when language of delivery was mentioned or could be inferred (local-language practitioner, translated manual, interpreter present, bilingual arm). Disagreements were resolved by discussion. Extraction fields: n, design, country, language of delivery, dose in sessions and minutes, delivery channel, guidance (self-guided, guided, interpreter-mediated), outcome measure, timepoints, and whether the L1 versus L2 contrast was within-study or between-study. Anything that could not be verified with a DOI or a stable URL was marked UNVERIFIED and excluded from any quantitative claim. One candidate Indian “regional-language helpline RCT” published in a multi-subject journal with no trial registration and no verifiable ethics record was marked UNVERIFIED and dropped.
Analysis plan, as pre-specified and as delivered	Pre-specified model: random-effects REML; Hedges g with 95 per cent CI; I^2, tau^2, 95 per cent prediction interval; Egger’s test and trim-and-fill; leave-one-out; RoB 2; GRADE. Pre-specified moderators: first versus second language; interpreter-mediated; machine-translated. Delivered analysis: narrative synthesis. A new pooled g from primary L1-versus-L2 trials was not computed. Between-study language-matching subgroups from prior meta-analyses are reported as the best available quantitative signal and are labelled as such. Interpreter-mediated studies are synthesised separately and are not pooled with the language-matching subgroups. Machine-translated delivery had no eligible controlled trial.

Risk of bias and certainty

RoB 2 was applied in spirit to any randomised trial that could have entered a pool. For the language-of-delivery contrast, almost every candidate was either non-randomised or randomised on a different factor. GRADE was applied to two claims: (1) first-language delivery outperforms second-language delivery; (2) culturally adapted programmes outperform unadapted programmes.

PRISMA 2020 flow (reconstructed from this search)

This flow is reconstructed from the searches run for this review. It is not the output of a registered software pipeline.

- Records identified: approximately 4,530 (PubMed targeted set after design filters and related-citation chaining ~1,100; PsycINFO ~900; Scopus ~1,400; Web of Science ~800; Google Scholar first 200; CTRI ~40; ClinicalTrials.gov ~60; OSF ~30).
- After de-duplication: approximately 2,800.
- Title and abstract excluded: approximately 2,650. Main reasons: child speech-and-language programmes; classroom language-of-instruction; aphasia; no wellbeing or symptom outcome; not adult.
- Full texts assessed: approximately 150.
- Full texts excluded: cultural adaptation without an isolable language contrast; child sample; no outcome; not a psychological or wellbeing programme; pre-2000; no controlled design; UNVERIFIED source.
- Included in the narrative synthesis: 6 prior meta-analyses as context; 5 interpreter or language-format primary studies in adults; 4 Indian or South-Asian programmes delivered in a local or preferred language but not randomised on language.
- Eligible for a new pooled L1-versus-L2 meta-analysis of the same programme: 0.

06 Results

Study characteristics

No eligible randomised trial assigned multilingual adults to the same wellbeing programme in a first language versus a second language.

What the field actually contains is three stacked layers.

Layer 1 — prior meta-analyses in which language is a moderator. These are between-study comparisons. A study delivered in Spanish to Latinx adults is compared with a study delivered in English to another group. Language travels with ethnicity, acculturation, measure translation and content adaptation.

Layer 2 — interpreter-mediated versus same-language delivery. These are closer to a language-of-delivery contrast, but they compare a three-person conversation with a two-person conversation. They are mostly not randomised.

Layer 3 — programmes delivered in a local language in India and other multilingual settings, with no L2 arm. These show that first-language delivery is feasible and often effective against usual care. They do not test language.

Table 1. Prior meta-analyses that speak to language of delivery

Source	k (N)	Language contrast	Main finding	Language-specific number
Griner & Smith 2006	76 (25,225)	Between-study moderator	Adapted programmes d = 0.45 (0.36–0.53)	Native-language (if not English) d = 0.49 (0.39–0.58) vs no match described d = 0.21 (0.01–0.40); “twice as effective”; Q = 6.1, p = .05
Smith, Domenech Rodríguez & Bernal 2011	65 (8,620)	Between-study	Adapted programmes d = 0.46 (0.36–0.56)	Number of adaptations predicted larger effects; language among the coded adaptations
Benish, Quintana & Wampold 2011	Direct-comparison set	Adapted vs unadapted bona fide programme	d = 0.32 on primary functioning	Illness-myth adaptation was the only significant unique moderator; language match was not
Chowdhary et al. 2014	16 pooled of 20 reviewed	Adapted depression programmes, including LMIC	SMD −0.72 (−0.94 to −0.49)	Language, context and practitioner most common adaptations; 15 of 20 studies adapted language
Hall et al. 2016	78 (13,998)	Adapted vs other conditions; adapted vs unadapted same programme	g = 0.67 overall; g = 0.52 (0.15–0.90) vs unadapted same programme	Language (English vs other) did not significantly moderate the continuous symptom outcome
Soto et al. 2018	99	Update of the Smith line	d = 0.50; 0.35 after publication-bias correction	Preferred-language delivery d = 0.59 vs not reported d = 0.35, p = .03; translated measures d = 0.66 vs 0.28; preferred language survived multivariate control (β = 0.18)

Source	k (N)	Language contrast	Main finding	Language-specific number
Arundell et al. 2021	30 RCT comparisons vs active	Adapted vs non-adapted active programmes	g = −0.43 (−0.61 to −0.25)	Language-translation (therapist) vs waitlist g = −0.82; vs active g = −0.47. About half used a same-language or bilingual practitioner or an interpreter

Table 2. Forest-style display of between-study language-matching subgroups (not a new pool)

These rows are subgroups published inside prior meta-analyses. They are not a new random-effects model run on primary L1-versus-L2 trials. Weights are not assigned because we did not pool.

Contrast	g or d	95% CI	What the row is
Griner & Smith: native-language delivery (non-English)	d = 0.49	0.39 to 0.58	Between-study subgroup
Griner & Smith: no language match described	d = 0.21	0.01 to 0.40	Between-study subgroup
Implied difference	≈ 0.28	Contrast p = .05	Not a within-study g
Soto et al.: preferred-language delivery	d = 0.59	—	Between-study subgroup
Soto et al.: language not reported	d = 0.35	—	Between-study subgroup
Implied difference	≈ 0.24	p = .03	Not a within-study g
Soto et al.: translated measures	d = 0.66	—	Measure language, not coach language
Soto et al.: measures not translated	d = 0.28	—	Measure language
Hall et al.: adapted vs unadapted same programme	g = 0.52	0.15 to 0.90	Cultural adaptation as a bundle
Chowdhary et al.: adapted depression programmes	SMD −0.72	−0.94 to −0.49	Language bundled with other adaptations
Arundell et al.: language-translation vs waitlist	g = −0.82	−1.25 to −0.40	Therapist language translation
Arundell et al.: language-translation vs active	g = −0.47	−0.73 to −0.20	Therapist language translation
Benish et al.: adapted vs bona fide unadapted	d = 0.32	—	Language not the unique moderator

Table 3. Moderator table — pre-specified contrasts

Moderator	Eligible controlled tests	Direction	Certainty
First vs second language, same programme, randomised	0 adult trials	Cannot estimate	Very low
First vs second language, between-study	2 major subgroups (Griner; Soto)	L1 larger by ~0.24 to 0.28	Low
Interpreter-mediated vs same-language practitioner	5 adult comparisons, 1 large	Mixed; 3 similar, 1 worse with interpreter, 1 both large	Very low
Machine-translated vs human L1	0 eligible trials	Unknown	—
Measure translation	Soto 2018 subgroup	Translated measures d = 0.66 vs 0.28	Low (still between-study)

Table 4. Interpreter-mediated and language-format primary studies in adults

Study	n	Design	Contrast	Result
Schulz et al. 2006	53	Non-randomised; refugees, US	Same-language practitioner vs interpreter	Both large effects; no significant difference
d'Ardenne et al. 2007	Three groups	Routine-care comparison; East London	Refugees with interpreter vs without vs English speakers	Refugees with and without interpreter did not differ; all groups improved
Brune et al. 2011	190	Retrospective; refugees, Germany	Interpreter vs no interpreter	Equivalent outcomes despite tougher baseline in the interpreter group
Sander et al. 2019	825	Retrospective cohort; Denmark	Interpreter vs no interpreter, 16 CBT-informed sessions	Interpreter arm improved less on the primary trauma measure and most secondaries
Klein Schiphorst, Levi & Manley 2022	133	Pre-post, non-equivalent groups; UK	Same-language counsellor vs interpreter	No significant difference on mood and worry scores; satisfaction high in both arms

Table 5. Indian and South-Asian programmes delivered in a local or preferred language — not L1-versus-L2 trials

Study	n / design	Languages	What was randomised	Why it cannot enter the L1–L2 pool
Patel et al. 2010, MANAS, Goa	Cluster RCT, primary care	Konkani, Marathi, Hindi or English; ~98–99% Konkani	Lay-counsellor collaborative care vs usual care	Language was an eligibility filter, not an arm
Patel et al. 2017, HAP, Goa	495, RCT	Local languages; Konkani-validated symptom measure	HAP plus enhanced usual care vs enhanced usual care	Delivered in L1. Not tested against an English HAP. Remission ratio 1.61; adjusted mean difference −7.57
Weobong et al. 2017, HAP 12-month	Same cohort	Same	Same	Gains held at 12 months; still no language arm
Pathare / Joag et al. 2023, Atmiyata, Gujarat	1,191, stepped-wedge cluster RCT	Gujarati community programme	Immediate vs delayed roll-out	Local-language delivery vs later local-language delivery. Recovery OR 2.2 at 3 months
Husain et al. 2024, ROSHNI-2	Multi-site UK RCT	Preferred languages: Urdu, Punjabi, Gujarati, Bengali, Tamil	Culturally adapted group programme vs usual care	Preferred-language delivery in the UK, not an India L1-versus-L2 trial. Recovery OR 1.97 at 4 months

Patel and colleagues' Healthy Activity Programme remains the strongest Indian demonstration that a brief, lay-delivered behavioural-activation programme can work in primary care when it is spoken in the language of the clinic. It is not a language trial.

One further adult trial is easy to misread. Lopez-Maya, Olmstead and Irwin (2019) randomised 76 Los Angeles adults, stratified into Spanish-speaking (n = 36) and English-speaking (n = 40) groups, to six weeks of mindful awareness practices or health education. The mindfulness advantage was d = 0.28 overall, 0.29 in the Spanish-format group and 0.30 in the English-format group. That is evidence that a programme delivered in the language the person actually uses can work in either language. It is not a head-to-head of L1 versus L2. People were not assigned against their preferred language.

Heterogeneity

A new I^2 was not computed. Heterogeneity in the parent meta-analyses was substantial. Soto and colleagues reported high between-study variance and a trimmed overall effect of d = 0.35 once small studies were accounted for. Hall and colleagues reported that language as a moderator was inconsistent across earlier reviews. Arundell and colleagues reported I^2 above 80 per cent in several language-translation subgroups. The honest reading is that the between-study L1 signal sits inside a noisy family of cultural-adaptation effects.

Publication bias

Soto and colleagues is the cleanest published correction: overall d = 0.50 before correction, d = 0.35 after. Griner and Smith's "twice as effective" sentence was not trimmed in the original paper. Any reader who quotes 0.49 versus 0.21 should keep the Soto correction in the same paragraph. We did not run a new Egger test.

Risk-of-bias summary

For the question this paper asks — does L1 beat L2 — risk of bias is high. The contrast is almost never randomised. People who receive a programme in their first language also tend to receive a culturally matched coach, translated measures, and content rewritten for their group. Confounding is structural. For the parent question — do culturally adapted programmes beat unadapted ones — risk of bias is mixed and has been graded by the source reviews.

Table 6. Risk-of-bias judgements for the language-of-delivery contrast

Study	Language randomised?	Overall for L1–L2 contrast
Griner / Soto language subgroups	No (between-study)	High
Hall 2016 language moderator	No within-study assignment	High for language; lower for adaptation
Klein Schiphorst 2022	No	High
Sander 2019	No; severity likely confounded	High
d'Ardenne 2007; Brune 2011; Schulz 2006	No	High
Patel 2017 HAP	Randomised on programme vs usual care, not language	Low for HAP vs usual care; not applicable for L1 vs L2
Lopez-Maya 2019	Randomised on programme vs education, stratified by language	Low for mindfulness vs education; not applicable for L1 vs L2

GRADE table *Table 7. Certainty of evidence*

Claim	Effect	Studies	Certainty
First-language delivery outperforms second-language delivery of the same programme	Between-study Δ ≈ 0.24 to 0.28	0 head-to-head RCTs; 2 meta-analytic subgroups	Very low (bias, inconsistency, indirectness, imprecision, suspected publication bias)
Culturally adapted programmes outperform unadapted versions of the same programme	g = 0.52 (Hall); d = 0.32 (Benish bona fide)	78 studies and the Benish direct-comparison set	Low to moderate
Interpreter-mediated delivery is worse than same-language delivery	Mixed	5 adult comparisons	Very low
Machine-translated delivery matches human L1 delivery	No estimate	0 trials	No evidence
An Indian working-adult L1 advantage exists	No estimate	0 Indian head-to-head trials	Very low

07 Discussion

PULL QUOTE

Four interpreter-mediated comparisons exist; they do not agree.

What the number means

The number that will be quoted is g ≈ 0.28, or “twice as effective.” Both phrases come from a between-study contrast published in 2006 and echoed, with a slightly smaller gap, in 2018. They do not come from a trial that handed the same workbook to one group in Hindi and to another group in English and watched what happened to sleep, worry and the Friday-evening slump.

Put plainly, 0.28 is a small-to-moderate extra push. Picture where it lands. A project manager in Chennai writes down the thought that kept him awake. In office English it becomes “some discomfort around family expectations.” In Tamil it is the insult that actually stung, and only that version still has heat. Naming shame as shame is what lets him choose a different next step; that choice needs the feeling in its real clothes. Noticing “I am having the thought that I am a bad son” as just a thought works when the words are close enough to bite. For him, the words that bite are Tamil.

In machine-learning language, this is a tokenizer problem. An English-trained model can ingest a Hindi sentence through translation, but the embedding space was built on a different corpus. Transfer learning helps. It is not the same as training on the language the user thinks in. In entrepreneurship language,

this is product-market fit versus a localisation toggle. Shipping 23 languages is a bet that the toggle is not enough — that the first screen has to speak the language in which the customer scolds herself.

The generous reading: the town-hall message in English is gone by the lift, while the same line from a manager in Kannada is repeated at home. One office language is efficient on a slide and expensive in the body. The cautious reading asks what else changed.

Benish, Quintana and Wampold give the caution its sharpest empirical form. When they compared culturally adapted programmes with bona fide unadapted programmes, the advantage shrank to d = 0.32, and the only unique moderator was whether the programme had rewritten the “illness myth” — the story of what is wrong and what will help. Language match, on its own, did not carry the model. That finding does not kill the L1 hypothesis. It says language is rarely the only moving part.

Hall and colleagues add a second caution. Their overall adaptation advantage was real (g = 0.67 against mixed controls; g = 0.52 against the unadapted twin). Language as a moderator of the continuous outcome was not reliable. A reviewer who stops at Griner’s “twice as effective” sentence is quoting a 2006 poster, not the 2016 test.

Soto and colleagues add a third caution and a second number. After publication-bias correction the overall adaptation effect fell from 0.50 to 0.35. Preferred-language delivery still looked larger (d = 0.59 versus 0.35). Translated measures looked larger still (d = 0.66 versus 0.28). If only one operational change is affordable, translating the measures may be as important as translating the coach. A person filling an English symptom form in a Hindi session is doing two jobs at once.

The interpreter studies refuse a slogan. Klein Schiphorst and colleagues found no outcome difference and high satisfaction in both arms. d’Ardenne and Brune found similar patterns in refugee clinics. Sander and colleagues, with 825 people, found smaller trauma-symptom gains when an interpreter sat in the room. Severity and language proficiency almost certainly confound that cohort. An interpreter is not automatically second-best. An interpreter is also not automatically equal. The data are not yet a policy.

Limitations

This review has the limitations of its evidence and the limitations of its method.

First, we did not pool, because the pre-specified floor of five eligible trials was not met. Readers who wanted a new forest plot of L1-versus-L2 RCTs will not find one. That absence is the result.

Second, the between-study numbers we do quote confound language with culture, acculturation, socioeconomic status, ethnicity matching and measure translation. Griner and Smith said so themselves. Their coding of “language matching” was an approximation. Studies that did not mention language may still have matched it. Studies that mentioned it may not have delivered it.

Third, the samples behind those numbers are mostly ethnic-minority adults in the United States and, in the interpreter set, trauma-affected refugees in Europe. Indian working adults in Bangalore, Pune, Hyderabad, Kochi and Mumbai are not those samples. Generalising from a Latinx clinic in Los Angeles to a WhatsApp thread in Indore is an act of hope, not a GRADE upgrade.

Fourth, this review was not prospectively registered. The PRISMA counts are reconstructed from the searches we ran. A registered update with dual independent extraction and a locked protocol would be stronger.

Fifth, we excluded unverifiable work. That is a feature. It also means a genuine small trial sitting in a regional Indian journal without a DOI may have been missed. We would rather miss a real paper than pool a fake one.

Sixth, machine-translated delivery — the thing every product team is about to ship — has no eligible controlled trial in this literature. The gap is not a licence.

What this means for an Indian workplace programme

WHAT THIS MEANS

What this means for an Indian workplace programme

An Indian workplace already runs a daily language experiment. The morning stand-up is in English. The grumble about appraisals in the lunch queue is in Bengali. A coaching programme that only speaks English is asking the professional to do the rep in the office dialect and the feeling in the home dialect at the same time. That is extra cognitive load. It is what turns a two-minute practice into homework: saved for later, then for never.

The between-study signal says first-language delivery is associated with a larger effect, on the order of g = 0.28. The GRADE rating says we must not pretend that number is a causal estimate for Indian employees. The Census says English is the mother tongue of 0.02 per cent of the country. The low-regret design is therefore obvious. Offer the rep in the language the person thinks in: if the Pune professional texts her sister in Marathi, her evening rep should arrive in Marathi. Keep an English track for the people who want it. Do not force the rest through a second-language filter in the name of a global brand voice.

WhatsApp-first delivery helps because that is where code-switching already lives: the family group in Malayalam, the project group in English, the same thumb. A 90-second breathing rep does not need a lecture. It needs a prompt the thumb will open between the local train and the office lift. A gratitude rep lands when it is written in the language the other person feels. A values exercise dies if “I want my father to be proud of me” has to be rewritten as “demonstrates ownership.” Build the programme as countable reps, in 23 languages,

with a human hand-off when the model runs out of road. Then run the missing trial: same reps, Hindi versus English, Tamil versus English, randomised, working adults, 8 weeks, wellbeing and symptom outcomes, India-resident data. Until that trial exists, 23 languages is not over-engineering. It is humility about what we do not know, paid for in the language people already speak.

Research gaps, naming the India gap

The India gap is not that India lacks wellbeing trials. India has HAP, MANAS, Atmiyata, SMART Mental Health, and a growing set of preferred-language programmes for South Asian adults in the UK. The India gap is narrower and more awkward. No verified Indian trial has taken one programme and randomised the language of delivery.

That trial is buildable. The population is sitting in every large employer's bilingual workforce. The intervention already exists in draft form — behavioural activation, gratitude, mindfulness, values, doing the opposite of what the mood demands, each cut into countable reps. The comparator is the same script in English. The outcome measures already have Indian-language versions for several scheduled languages. The missing pieces are registration, sample size, and the willingness to risk a null.

Other gaps follow. Machine translation versus bilingual human coaching has not been tested. Interpreter-mediated coaching in Indian languages has not been tested in working-adult samples. Dose — how many minutes of L1 are worth how many minutes of L2 — is unknown. Whether measure translation alone captures most of the Soto 0.66-versus-0.28 gap is unknown. Whether the L1 advantage is larger for people with lower English fluency, as acculturation patterns in the older US studies suggest, is unknown in India and is the moderator that would change procurement.

A null result on that future trial would still be publishable. A small L1 advantage would change how employee assistance programmes are bought. A large L1 advantage would change how every Indian coaching product is engineered. None of those sentences can be written honestly today.

08 FAQ

Q1 Does coaching in my mother tongue work better than English?

Between-study data say yes by about g = 0.28, or roughly twice the effect. No trial has randomly assigned the same programme to L1 versus L2 in Indian adults. The honest answer is: probably, and we do not yet have the clean experiment.

Q2 **How many Indian trials compared Hindi, Tamil or Marathi with English?**

Zero. Large Indian programmes such as the Healthy Activity Programme were delivered in local languages, but language itself was never the randomised factor. Feasibility is not the same as a head-to-head test.

Q3 **If I need an interpreter, is the programme wasted?**

Four adult comparisons exist. Three found similar outcomes with and without an interpreter; in one of them, a 133-person UK study, satisfaction was high in both arms. One 825-person cohort found smaller gains with an interpreter. An interpreter is not second-best by default.

Q4 **Can a machine-translated script replace a first-language coach?**

We found no eligible controlled trial of machine-translated psychological programmes in multilingual adults. That is an evidence gap, not a green light. Translation without a bilingual coach is a hypothesis.

Q5 **Why does EmotionApp offer 23 languages if the pooled trial does not exist?**

Because 0.02 per cent of Indians have English as a mother tongue, and the between-study signal already favours matching language. Offering the language is the low-regret move while the head-to-head trial is missing.

09

References

Arundell, L.-L., Barnett, P., Buckman, J. E. J., Saunders, R., & Pilling, S. (2021). The effectiveness of adapted psychological interventions for people from ethnic minority groups: A systematic review and conceptual typology. Clinical Psychology Review, 88, 102063. https://doi.org/10.1016/j.cpr.2021.102063

Benish, S. G., Quintana, S., & Wampold, B. E. (2011). Culturally adapted psychotherapy and the legitimacy of myth: A direct-comparison meta-analysis. Journal of Counseling Psychology, 58(3), 279–289. https://doi.org/10.1037/a0023626

Brune, M., Eiroá-Orosa, F. J., Fischer-Ortman, J., Delijaj, B., & Haasen, C. (2011). Intermediated communication by interpreters in psychotherapy with traumatized refugees. International Journal of Culture and Mental Health, 4(2), 144–151. https://doi.org/10.1080/17542863.2010.537821

Chowdhary, N., Jotheeswaran, A. T., Nadkarni, A., Hollon, S. D., King, M., Jordans, M. J. D., Rahman, A., Verdeli, H., Araya, R., & Patel, V. (2014). The methods and outcomes of cultural adaptations of psychological treatments for depressive disorders: A systematic review. Psychological Medicine, 44(6), 1131–1146. https://doi.org/10.1017/S0033291713001785

d'Ardenne, P., Ruaro, L., Cestari, L., Fakhoury, W., & Priebe, S. (2007). Does interpreter-mediated CBT with traumatized refugee people work? A comparison of patient outcomes in East London. Behavioural and Cognitive Psychotherapy, 35(3), 293–301. https://doi.org/10.1017/S1352465807003645

Griner, D., & Smith, T. B. (2006). Culturally adapted mental health intervention: A meta-analytic review. Psychotherapy: Theory, Research, Practice, Training, 43(4), 531–548. https://doi.org/10.1037/0033-3204.43.4.531

Hall, G. C. N., Ibaraki, A. Y., Huang, E. R., Marti, C. N., & Stice, E. (2016). A meta-analysis of cultural adaptations of psychological interventions. Behavior Therapy, 47(6), 993–1014. https://doi.org/10.1016/j.beth.2016.09.005

Husain, N., et al. (2024). Efficacy of a culturally adapted, cognitive behavioural therapy-based intervention for postnatal depression in British South Asian women (ROSHNI-2). The Lancet. https://doi.org/10.1016/S0140-6736(24)01612-X

Klein Schiphorst, A. T., Levi, N., & Manley, J. (2022). Found in translation? The effectiveness of counselling using an interpreter, compared to counselling with a same language counsellor, in a non-English speaking population. International Journal for the Advancement of Counselling, 44, 628–645. https://doi.org/10.1007/s10447-022-09475-z

Lopez-Maya, E., Olmstead, R., & Irwin, M. R. (2019). Mindfulness meditation and improvement in depressive symptoms among Spanish- and English speaking adults: A randomized, controlled, comparative efficacy trial. PLOS ONE, 14(7), e0219425. https://doi.org/10.1371/journal.pone.0219425

Office of the Registrar General & Census Commissioner, India. (2011). Census of India 2011: Data on language and mother tongue. https://censusindia.gov.in/

Patel, V., Weiss, H. A., Chowdhary, N., Naik, S., Pednekar, S., Chatterjee, S., De Silva, M. J., Bhat, B., Araya, R., King, M., Simon, G., Verdeli, H., & Kirkwood, B. R. (2010). Effectiveness of an intervention led by lay health counsellors for depressive and anxiety disorders in primary care in Goa, India (MANAS): A cluster randomised controlled trial. The Lancet, 376(9758), 2086–2095. https://doi.org/10.1016/S0140-6736(10)61508-5

Patel, V., Weobong, B., Weiss, H. A., Anand, A., Bhat, B., Katti, B., Dimidjian, S., Araya, R., Hollon, S. D., King, M., Vijayakumar, L., Park, A. L., McDaid, D., Wilson, T., Velleman, R., Kirkwood, B. R., & Fairburn, C. G. (2017). The Healthy Activity Program (HAP), a lay counsellor-delivered brief psychological treatment for severe depression, in primary care in India: A randomised controlled trial. The Lancet, 389(10065), 176–185. https://doi.org/10.1016/S0140-6736(16)31589-6

Pathare, S., Joag, K., Kalha, J., Pandit, D., Krishnamoorthy, S., Chauhan, A., et al. (2023). Atmiyata, a community champion led psychosocial intervention for common mental disorders: A stepped wedge cluster randomized controlled trial in rural Gujarat, India. PLOS ONE, 18(6), e0285385. https://doi.org/10.1371/journal.pone.0285385

Sander, R., Laugesen, H., Skammeritz, S., Mortensen, E. L., & Carlsson, J. (2019). Interpreter-mediated psychotherapy with trauma-affected refugees: A retrospective cohort study. Psychiatry Research, 271, 684–692. https://doi.org/10.1016/j.psychres.2018.12.058

Schulz, P. M., Resick, P. A., Huber, L. C., & Griffin, M. G. (2006). The effectiveness of cognitive processing therapy for PTSD with refugees in a community setting. Cognitive and Behavioral Practice, 13(4), 322–331. https://doi.org/10.1016/j.cbpra.2006.04.011

Smith, T. B., Domenech Rodríguez, M., & Bernal, G. (2011). Culture. Journal of Clinical Psychology, 67(2), 166–175. https://doi.org/10.1002/jclp.20757

Smith, T. B., & Trimble, J. E. (2016). Culturally adapted mental health services: An updated meta-analysis of client outcomes. In T. B. Smith & J. E. Trimble, Foundations of multicultural psychology: Research to inform effective practice (pp. 129–144). American Psychological Association. https://doi.org/10.1037/14733-007

Soto, A., Smith, T. B., Griner, D., Domenech Rodríguez, M., & Bernal, G. (2018). Cultural adaptations and therapist multicultural competence: Two meta-analytic reviews. Journal of Clinical Psychology, 74(11), 1907–1923. https://doi.org/10.1002/jclp.22679

Weobong, B., Weiss, H. A., McDaid, D., Singla, D. R., Hollon, S. D., Nadkarni, A., Park, A. L., Bhat, B., Katti, B., Anand, A., Dimidjian, S., Araya, R., King, M., Vijayakumar, L., Wilson, G. T., Velleman, R., Kirkwood, B. R., Fairburn, C. G., & Patel, V. (2017). Sustained effectiveness and cost-effectiveness of the Healthy Activity Programme, a brief psychological treatment for depression delivered by lay counsellors in primary care: 12-month follow-up of a randomised controlled trial. PLOS Medicine, 14(9), e1002385. https://doi.org/10.1371/journal.pmed.1002385

About the author

ABOUT THE AUTHOR

Dr Atul Aswani; Mumbai Psychiatrist and trained as a Results Coach
Dr Atul Aswani is a Mumbai Psychiatrist and trained as a Results Coach. He founded EmotionApp, a non-clinical emotional-fitness platform for Indian professionals. Practices are delivered as countable daily reps, with an AI coach and human hand-off, on WhatsApp first, in 23 Indian languages. Data stay in India and follow DPDP. The work sits under ISO 9001 and ISO 13485.

HOW TO CITE

Aswani A. Mother Tongue, Bigger Effect?: A Meta-Analysis of Language-of-Delivery in Psychological Interventions. EmotionApp Research Paper 39. Mumbai: Tranquilo Meditek Sytems Private Limited; 2026. https://emotionapp.in/research/language-of-delivery-meta-analysis

LAST UPDATED

22 September 2026

NOTE

EmotionApp is a non-clinical emotional-fitness coaching platform. This paper describes practices, techniques and coaching. It is not therapy, treatment, diagnosis or medical advice.

SAFETY NOTE

If you are in distress, please contact a qualified professional or Tele-MANAS on 14416.