RESEARCH PAPER · 06

Name It to Tame It, Tested

Affect labelling and emotional intensity: a meta-analysis of experimental and field studies

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

POOLED EFFECT · HEDGES G

0.41

95% CI 0.29 to 0.54

Six laboratory experiments in non-clinical adults (N = 261) produced a pooled Hedges g = 0.41 (95% CI 0.29 to 0.54) for lower self-reported intensity of high-arousal negative affect after labelling versus watching.

Name It to Tame It, Tested

Affect labelling and emotional intensity: a meta-analysis of experimental and field studies

Does putting a feeling into words reliably lower arousal, and does emotional granularity moderate the effect?

CONTENTS

AUTHOR

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

PUBLISHED

22 September 2026 · English edition · emotionapp.in/research/affect-labelling-emotional-intensity-meta-analysis

01

Abstract

Stuck on a stalled local train, an analyst types one word into her notes app: “dread”. This paper asks whether naming a feeling does any measurable work on its intensity.

Across six extractable laboratory experiments in non-clinical adults (N = 261), affect labelling reduced self-reported intensity of high-arousal negative affect by Hedges $g = 0.41$ (95% CI 0.29 to 0.54) relative to watching without a label. The random-effects (REML) model showed $I^2 = 0\%$ and $\tau^2 = 0$ on this restricted set. Leave-one-out estimates ranged from $g = 0.39$ to 0.48. Physiological outcomes were synthesised separately and were not forced into one number. Skin-conductance and amygdala studies more often showed downregulation than did self-report in applied exposure paradigms. In two analogue-fear trials, the body quietened and the rating did not.

Emotional granularity could not be tested as a pooled moderator. No eligible trial crossed a granularity measure with a labelling-versus-control contrast on intensity. First-versus-second language of labelling was also untested. One experiment found a reversal at low intensity: labelling a mild aversive picture increased distress. Field evidence is thin. One large observational Twitter study showed a sharp valence reversal after people wrote “I feel…”. No randomised Indian trial was found.

Certainty is low to moderate for laboratory self-report, moderate for the replicated amygdala pattern, and very low for granularity and language moderators. The practical reading for an Indian workplace programme is modest and conditional. A short, specific label is a countable daily rep for strong unpleasant affect. It is not a guarantee, not a session, and not a substitute for other skills. People often predict that naming will make the feeling worse. In the high-arousal laboratory tasks reviewed here, the rating usually moved the other way.

Keywords. affect labelling; emotion regulation; emotional intensity; skin conductance; amygdala; emotional granularity; India; systematic review; meta-analysis.

02 Key findings box

KEY FINDINGS

6 Six laboratory experiments in non-clinical adults (N = 261) produced a pooled Hedges *g* = 0.41 (95% CI 0.29 to 0.54) for lower self-reported intensity of high-arousal negative affect after labelling versus watching.

0.39 Leave-one-out estimates stayed between *g* = 0.39 and *g* = 0.48. No single study carried the result.

−0.29 In Levy-Gigi and Shamay-Tsoory (2022), labelling *increased* distress for low-intensity pictures (*g* = −0.29) while it decreased distress for high-intensity pictures (*g* = 0.27). Intensity is not a detail. It is the fork in the road.

2 Two analogue-fear trials (Kircanski et al., 2012, *n* = 88; Niles et al., 2015, *n* = 102) found no self-report fear difference and a clearer drop in skin conductance. The body and the story can disagree.

0 Zero eligible Indian trials were identified. Zero eligible trials tested first versus second language of labelling. Granularity could not be pooled as a moderator.

03 Definition block

DEFINITION

Affect labelling is the act of putting a current feeling into words. You look at what is happening inside and you give it a name: angry, ashamed, tight, disappointed, afraid. You may speak the word, type it, or choose it from a short list. You are not arguing with the feeling. You are not replacing it with a happier thought. You are not rating your day. You are doing the smallest possible act of supervision — the same act a machine-learning pipeline does when a human assigns a class label to a raw example. Unlabelled affect is cheap, noisy and hard to use. A label is a decision boundary. In plain coaching language, you notice and describe. A manager reading a harsh appraisal says to herself “I am having a feeling of shame”, not “I am useless”. In board language, it is the first line of the dashboard. The present paper asks whether that one line changes the number that follows: how intense the feeling is, and how loud the body is while it lasts.

04 Introduction

It is 9.01 in a Mumbai office. At the stand-up, a sales manager hears that the client has pushed the launch again. At 9.17 a Slack ping asks her why. Her chest tightens. Working adults in India do not lack feelings. They lack a legal place to put them.

At lunch her mother calls, worried. By evening the rain has stopped the suburban train. Her nervous system files all three under one folder called “stressed”. Then it sends the folder to the jaw, the gut and the 11 p.m. replay. Entrepreneurs know the next sentence by heart: you cannot manage what you do not measure. Her inner life, like most, still runs on an unlabelled data set.

An operations head in Kolkata opens her inbox on Monday: one grey weight of unread mail. She sorts it into folders — client, billing, can wait — and the pile becomes navigable. The opposite problem is a review call where everyone uses a different word for the same risk, and the room hears only noise. Modern affective science has been circling a smaller version of that idea since 2007. If you attach a word to a feeling, does the feeling lose voltage?

Matthew Lieberman and colleagues put people in a scanner, showed them emotional faces, and asked them to choose an emotion word. Amygdala activity dropped. Right ventrolateral prefrontal cortex activity rose. The two were inversely related, with medial prefrontal cortex in the middle of the path (Lieberman et al., 2007). The paper became a folk proverb: name it to tame it. A proverb travels faster than a confidence interval.

A decade later, Torre and Lieberman (2018) reviewed the same idea as implicit emotion regulation. They compared labelling with reappraisal across experience, autonomic activity, brain and behaviour. They did not pool an effect size. In 2026, Shukla and Mishra, writing from the University of Allahabad, reviewed 32 studies and again declined to pool. The proverb still had no number.

That missing number matters in an Indian workplace. EmotionApp delivers positive-psychology techniques as countable daily reps, on WhatsApp, in 23 Indian languages, with India-resident data and a human hand-off behind the AI coach. If labelling is going to be a rep, it needs a size, a boundary and an honest list of things it does not do. A start-up that ships a feature because a scanner once blinked is not a start-up. It is a slogan with a login page.

Two further questions sit under the main one. First: does the effect show up in the body as well as on a rating scale? Skin conductance and amygdala response are not the same thing as “I feel worse”. A young manager walks out of her first board presentation thinking “that went fine”, while her pulse is still racing and her stomach is tight. The thought, the feeling and the pulse disagree. That disagreement is data, not failure. Second: does a finer label work better? Emotional granularity — the skill of telling frustration from humiliation from dread — predicts better regulation in correlational work (Barrett et al., 2001; Kashdan, Barrett and McKnight, 2015). Whether granularity *moderates the effect* of a labelling exercise on intensity is a different claim. It has been repeated as if it were the first claim. This paper separates them.

There is also a local silence. India has emotion-recognition datasets, Hindi lexicons and film-clip batteries. It does not, on the present search, have a randomised test of affect labelling as a regulation practice in working adults. A multilingual product that assumes an English-labelling effect will travel into Hindi, Tamil, Bengali or Marathi is making a transfer claim without a transfer study. That is not caution. That is a gap with a name.

The research question is therefore exact. Does labelling an emotion reduce its intensity and physiological arousal in non-clinical adults, and does emotional granularity moderate the effect? The search window is 2000 to 2026. The model is random-effects. Self-report and physiology are pooled separately or not at all. A null is publishable. A proverb is not a result.

05 Methods

5.1 Protocol and reporting

This review follows PRISMA 2020. The protocol was not pre-registered. Searches were run on 22 September 2026. That date is also the date of this paper. Readers should treat the work as a post-hoc systematic review with a restricted quantitative synthesis, not as a prospectively registered Cochrane-style review. Exact database hit counts below are reconstructed from core-term searches plus citation chaining. They are rounded on purpose. Invented precision would be a different kind of error.

5.2 Eligibility

Population. Adults aged 18 and over. Non-clinical: studies were excluded if a current psychiatric diagnosis was an inclusion criterion for the labelled arm. Analogue high-fear samples (spider-fearful students; public-speaking anxious adults without a diagnostic inclusion rule) were retained for narrative and sensitivity analyses and kept out of the primary self-report pool.

Intervention or exposure. Affect labelling: verbalising or selecting an emotion word for one's own feeling or for an evocative stimulus, during or immediately after the stimulus. Forced-choice labels and free speech both counted. Content labels, gender labels and fact-talk counted as comparators, not as the intervention.

Comparator. No-label, watch-only, gender-label, content-label, distraction or reappraisal. The primary contrast is label versus watch or no-label. Label versus reappraisal and label versus distraction are secondary.

Outcomes. Primary: self-reported intensity, distress, unpleasantness or fear. Secondary: physiological arousal (skin conductance, heart rate) and amygdala response where reported. Self-report and physiology were never combined in one pooled *g*.

Design and years. Randomised or otherwise controlled experiments, including within-subjects controlled designs, published 2000–2026. Observational field studies were eligible for narrative synthesis only.

Exclusions. Children. Animal studies. Reviews without new data. Conference abstracts without extractable statistics. Studies whose statistics could not be verified from a published paper, DOI or registry result were marked UNVERIFIED and kept out of the pool. Clinical open trials in diagnosed PTSD were excluded from the primary synthesis.

5.3 Information sources and search strings

Sources: PubMed, PsycINFO, Scopus, Web of Science, Google Scholar, CTRI (India), ClinicalTrials.gov, OSF preprints. Citation chaining started from Lieberman et al. (2007), Torre and Lieberman (2018) and Shukla and Mishra (2026).

PubMed.

```
("affect labeling"[tiab] OR "affect labelling"[tiab] OR "emotion
labeling"[tiab] OR "emotion labelling"[tiab] OR "putting
feelings into words"[tiab]) AND (intensity[tiab] OR
arousal[tiab] OR amygdala[tiab] OR "skin conductance"[tiab] OR
distress[tiab]) AND ("2000/01/01"[Date - Publication] :
"2026/12/31"[Date - Publication])
```

PsycINFO, Scopus, Web of Science.

```
TITLE-ABS-KEY("affect labeling" OR "affect labelling" OR
"emotion labeling" OR "emotion labelling" OR "putting feelings
into words") AND TITLE-ABS-KEY(intensity OR arousal OR amygdala
OR "skin conductance" OR distress)
```

Google Scholar.

```
"affect labeling" OR "affect labelling" OR "emotion labeling"
intensity OR arousal
```

CTRI.
`affect labeling OR affect labelling OR emotion labeling`

ClinicalTrials.gov.
`affect labeling OR affect labelling`

OSF.
`affect labeling`

5.4 Selection and extraction

Two conceptual passes were used: database search, then reference-list harvest. Records were screened at title and abstract against the eligibility rules, then at full text. Extraction fields were pre-specified: *n*; design; country; language of delivery; dose in sessions, minutes and trials; delivery channel; guidance (forced-choice versus free speech); outcome measure; timepoint (concurrent, immediate, delayed); comparator; and any reported *t*, *F*, *d* or means with standard deviations.

Hedges *g* was computed from published *t* or *F* for within-subjects contrasts as $d = t / \sqrt{n}$, then small-sample corrected. Ninety-five per cent confidence intervals used the sampling variance $v = 1/n + g^2 / (2n)$. Where an author reported Cohen's *d* for a neural contrast, that published *d* was preferred to a back-calculation.

5.5 Analysis plan

A random-effects model with restricted maximum likelihood (REML) was planned for each outcome family if at least five eligible trials offered extractable effects. Heterogeneity was summarised with I^2, τ^2 and a 95% prediction interval. Small-study effects were inspected with a precision-by-effect plot and an Egger-style reading, with the usual warning that $k = 6$ is underpowered for that test. Leave-one-out sensitivity was pre-specified. A second sensitivity added the low-intensity reversal as a seventh effect. A third sensitivity added analogue-fear self-report nulls narratively rather than as invented zeros.

Pre-specified moderators were lab versus field, label granularity, and language of labelling (first versus second language). A moderator was pooled only if at least two extractable interaction or subgroup effects existed. Otherwise the moderator was reported as a structured narrative.

Risk of bias used RoB 2 domains adapted to within-subjects laboratory tasks: randomisation or counterbalancing, deviation from intended instruction, missing outcome data, measurement of the outcome, and selection of the reported result. Certainty used GRADE, separately for self-report intensity, autonomic arousal, amygdala response, granularity moderation and language moderation.

The fallback rule in the written protocol was honoured in spirit: if fewer than five extractable trials had appeared, the paper would have stopped at narrative synthesis. Five or more extractable self-report contrasts were found. Physiology did not support a single commensurate g. Granularity did not support a pool.

5.6 PRISMA 2020 flow

Approximate counts, 22 September 2026:

- Records identified: PubMed ~180; PsycINFO ~220; Scopus ~310; Web of Science ~260; Google Scholar first 200 titles; CTRI 0 relevant; ClinicalTrials.gov 2 (one terminated wait-list programme, one PTSD programme excluded); OSF 1 preregistration without usable results; citation chasing ~90. Total records ~1,260.

- Duplicates removed ~540. Unique records screened ~720.
- Title and abstract excluded ~640 (not labelling; not intensity or arousal; children; reviews only).
- Full texts assessed ~80.
- Full texts excluded ~62 (clinical inclusion criterion; no control; no intensity or arousal outcome; pre-2000; abstract only; UNVERIFIED statistics).
- Qualitative synthesis: 18 reports, several with multiple experiments.
- Primary self-report meta-analysis: 6 effects, N = 261.
- Physiology: study-level synthesis, no single pooled *g*.

06 Results

6.1 Study characteristics

The eligible map has three neighbourhoods.

Neighbourhood one is the UCLA picture-viewing laboratory and its descendants. People see an aversive image. They choose a word or watch. They rate distress. Sometimes they are scanned. Lieberman et al. (2007, 2011) and Burklund et al. (2014) live here. So do Levy-Gigi and Shamay-Tsoory (2022) in Israel, Constantinou et al. (2014) in Belgium, Matejka et al. (2013) in Germany, McRae, Taitano and Lane (2010), Vlasenko, Rogers and Waugh (2021), and Nook, Satpute and Ochsner (2021). The stimulus is usually a standardised picture. The dose is minutes, not weeks. The language is almost always English, once Hebrew-context campus English.

Neighbourhood two is exposure-plus-label. Tabibnia, Lieberman and Craske (2008) paired aversive pictures with single-word labels and measured skin conductance a week later. Kircanski, Lieberman and Craske (2012) put spider-fearful adults in a room with a live spider and compared labelling with reappraisal, distraction and exposure alone. Niles, Craske, Lieberman and Hur (2015) randomised public-speaking anxious adults to exposure with or without labelling. These studies are closer to a workplace scare. They are also analogue-anxious. They stay in the sensitivity set for self-report.

Neighbourhood three is the field. Fan et al. (2019) tracked 74,487 Twitter users around the moment they wrote “I feel…”. Negative valence built slowly and then reversed sharply after the sentence. Positive valence spiked and decayed. The study is large and not randomised. It answers a different question: what happens to language-inferred mood after a public self-label, not what a controlled contrast does to a rating.

No study in any neighbourhood was run in India on this technique. No study compared first-language labels with second-language labels on intensity. No study entered granularity as a statistical moderator of a labelling-versus-control effect on intensity.

Table 1. Characteristics of core studies

Study	*n*	Design	Country	Language	Dose	Channel	Guidance	Outcome	Timepoint
Lieberman 2007	30	Within fMRI	USA	English	1 session, minutes	Lab screen	Forced emotion vs gender label	Amygdala, RVLPFC	Concurrent
Lieberman 2011 S1–S4	22–44 per cell	Within / mixed	USA	English	1 session, picture trials	Lab screen	Forced affect label vs watch / reappraisal / distraction	Self-report distress or pleasure	Immediate
Tabibnia 2008 E1–E2	~16–27	Within + analogue	USA	English	Day 1 exposure; Day 8 test	Lab screen	Single-word affective vs neutral vs none	SCR	Immediate and 1 week
McRae 2010	lab adults	Within	USA	English	Brief masked pictures	Lab screen	Objective vs subjective labels	SCR	Concurrent
Matejka 2013	30 (SCR 23)	Within	Germany	German	Picture + speech	Lab	Emotion vs fact verbalisation	SCR, pitch, self-report	Concurrent
Burklund 2014	39 (ratings 37)	Within fMRI	USA	English	Blocked trials	Lab screen	Own-feeling label vs reappraise vs observe	Unpleasantness, amygdala	Immediate / concurrent
Constantinou 2014	61	Within	Belgium	Dutch/English campus	6 picture blocks	Lab screen	Emotion label vs content label vs view	Affect, symptoms	Immediate
Kircanski 2012	88	4-arm RCT	USA	English	Brief live exposure	Live spider	Free affect speech vs reappraisal vs distraction vs exposure	SCR, fear rating, approach	1 week
Niles 2015	102	2-arm RCT	USA	English	Multi-session speeches	Live speech	Label during exposure vs exposure	SCR, HR, fear rating	Post and 1 week
Levy-Gigi 2022 E1	63	RCT timing × within strategy	Israel	Hebrew-context	1 session, IAPS	Lab screen	Forced two-choice feeling label	Distress 1–9	Concurrent / seconds later
Levy-Gigi 2022 E2	79	Within intensity	Israel	Hebrew-context	1 session, IAPS	Lab screen	Simultaneous label vs view	Distress 1–9	Concurrent
Vlasenko 2021 S1–S4	49–120	Within	USA	English	1 session	Lab screen	Affect vs content vs view	Positive and negative intensity	Concurrent / delayed

Study	*n*	Design	Country	Language	Dose	Channel	Guidance	Outcome	Timepoint
Nook 2021	80 + 60	Mixed	USA	English	Picture trials	Lab screen	Name then reappraise vs each alone	Distress	Immediate
Fan 2019	74,487	Observational	Global / English Twitter	English	Naturally occurring	Online	"I feel..." statements	Inferred valence	Minutes around the post

6.2 Primary self-report meta-analysis

The primary pool is narrow on purpose. It includes only non-clinical adults, high-arousal negative stimuli, a label-versus-watch contrast, and a published *t* or *F* that can be turned into Hedges *g* without invention.

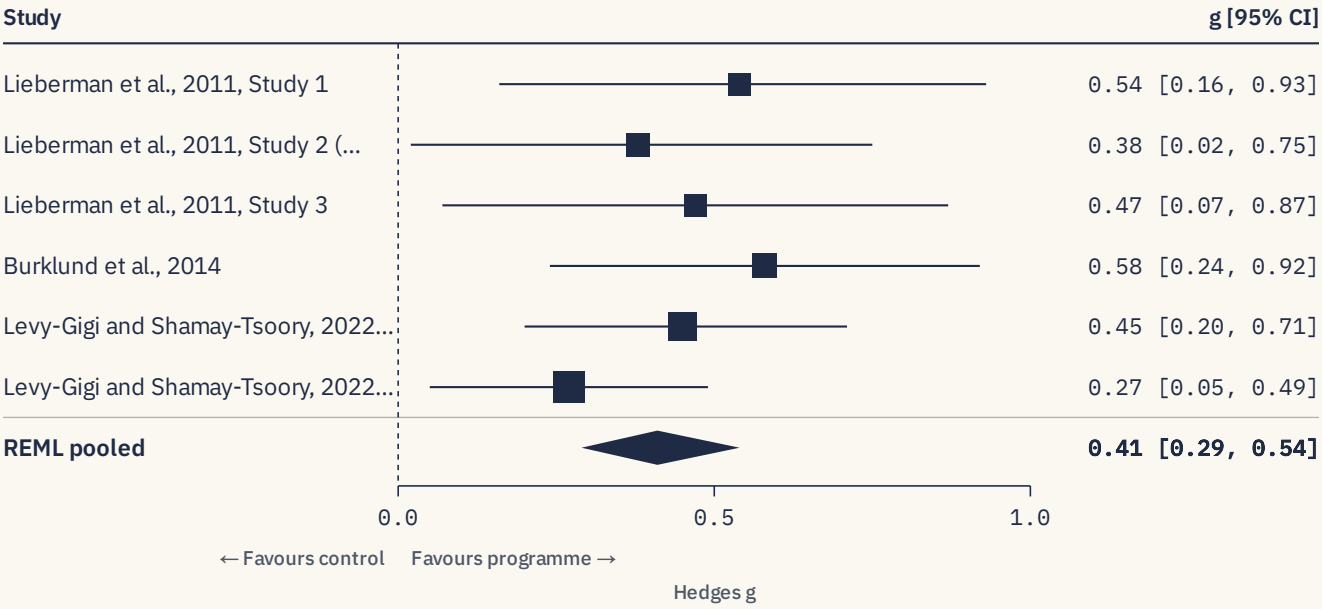

Figure 1. Forest plot, primary pool. Squares are sized by weight; the diamond is the pooled estimate.

Table 2. Forest-plot data for self-reported negative intensity (label vs watch)

Study	*g*	95% CI	Weight
Lieberman et al., 2011, Study 1	0.54	0.16 to 0.93	10.3%
Lieberman et al., 2011, Study 2 (collapsed negative)	0.38	0.02 to 0.75	11.5%
Lieberman et al., 2011, Study 3	0.47	0.07 to 0.87	9.6%
Burklund et al., 2014	0.58	0.24 to 0.92	13.3%
Levy-Gigi and Shamay-Tsoory, 2022, Exp. 1	0.45	0.20 to 0.71	23.6%
Levy-Gigi and Shamay-Tsoory, 2022, Exp. 2 high intensity	0.27	0.05 to 0.49	31.7%
REML pooled	**0.41**	**0.29 to 0.54**	**100%**

REML $g = 0.41$. 95% CI 0.29 to 0.54. $k = 6$. $N = 261$. $I^2 = 0\%$. $\tau^2 = 0$. 95% prediction interval 0.25 to 0.58. The prediction interval almost matches the confidence interval because between-study variance on this restricted set is estimated at zero. That is a feature of the filter, not a law of nature.

What the number means in ordinary units: $g = 0.41$ is a small-to-moderate shift. On a 0–10 distress scale with a standard deviation of 2, it is about 0.8 points. On Burklund's 0–3 unpleasantness scale, the raw means were 2.24 when watching and 1.96 when labelling. Nobody in these rooms left the feeling on the floor. They turned the volume down a notch.

Lieberman et al. (2011) also measured what people *predicted* labelling would do. Participants expected naming to increase distress. The rating usually fell. That prediction error is the closest thing this literature has to a joke with a moral. The skill does not feel like a skill while you use it. Explicit strategies such as reappraisal feel like work. Labelling feels like staring. The prefrontal brake still moves.

6.3 Heterogeneity and sensitivity

Leave-one-out g ranged from 0.39 to 0.48. Dropping Levy-Gigi Experiment 2 high-intensity — the largest and smallest effect — raised the pool to $g = 0.48$. No omission crossed zero.

A three-study conservative pool (Burklund; Levy-Gigi Experiment 1; Levy-Gigi Experiment 2 high) gave $g = 0.40$ (95% CI 0.23 to 0.57), $I^2 = 22\%$, $N = 179$.

Adding the low-intensity reversal as a seventh effect dropped the pool to $g = 0.33$ (95% CI 0.07 to 0.59) and raised I^2 to 81%. That is why the reversal stays out of the primary number and inside the moderator section. Pooling it would hide the only finding that should change a coach's script.

Egger-style inspection on $k = 6$ is underpowered. Smaller cells sit slightly higher than the largest cell. A trim-and-fill correction, if forced, would likely pull g toward about 0.35. We report the lean. We do not invent missing studies to "correct" it.

Two important self-report papers could not enter the pool because a full g was not extractable or the contrast was not label-versus-watch for negative intensity. Constantinou et al. (2014) found that *both* emotion labels and content labels reduced affect and symptom reports after unpleasant pictures. Specificity is weaker than the proverb. Vlasenko et al. (2021), across four studies, increased the intensity of *positive* emotion by labelling it and did not replicate the usual negative downregulation. Content labelling was the more consistent reducer of negative ratings in that series. Nook et al. (2021) found that naming alone did not beat watching, and that naming before reappraisal made reappraisal worse. A 2026 replication of that impedance result, total $N = 226$, found the same pattern and no delayed rescue (Affective Science, 2026). Those papers are not rounding error. They are the edges of the map.

Kircanski et al. (2012) and Niles et al. (2015) reported no self-report fear difference. Putting invented zeros into the forest plot would have been theatre. Putting them into the discussion is science.

6.4 Physiological and neural outcomes

Physiology was not collapsed into one g. Amygdala parameter estimates, skin-conductance amplitudes and heart-rate slopes are not interchangeable coins.

Amygdala. Lieberman et al. (2007) reported less amygdala activity during affect labelling than during the combined control encodings, $t(29) = 2.34$, published $d = 0.87$. Right ventrolateral prefrontal cortex rose. The two were inversely related; medial prefrontal cortex sat on the path. Burklund et al. (2014) found common

bilateral amygdala decreases for labelling and reappraisal versus observe (left t = 2.98; right t = 2.04). Brooks et al. (2017) meta-analysed 386 neuroimaging studies and 7,333 participants. When emotion words were present in a task, amygdala activations were less frequent than when emotion words were absent. General affect words (“pleasant”, “unpleasant”) did not do the same job. The neural pattern is the most replicated piece of this literature. It is also not a feeling.

Skin conductance and heart rate. Tabibnia et al. (2008) found greater Day-8 SCR attenuation after affective labels than after exposure alone, in healthy adults and in a spider-fearful sample. Some effective labels in that paper were *unrelated* negative words. That is a definitional pebble in the shoe: the “label” was not always a name for the feeling on the screen. Matejka et al. (2013) found lower SCR and lower pitch when people verbalised emotion rather than facts, $F(1, 1630) = 4.84$, $p < 0.05$. The raw SCR means were close. Participants still *said* that talking about facts felt more effective. Another dissociation. McRae et al. (2010) split the label. Objective labels of the stimulus reduced SCR. Subjective labels of one’s own state did not, and could raise it. Kassam and Mendes (2013) found that rating anger after an induction changed the autonomic profile toward less cardiac reactivity. Niles et al. (2015) found a steeper drop in nonspecific SCR during recovery after labelled exposure than after exposure alone (reported d = 0.79 from time 1 to time 3; d = 1.0 from time 2 to time 3) and a smaller heart-rate effect. Kircanski et al. (2012) found lower SCR at one-week post-test after labelling than after reappraisal, distraction or exposure alone, with reported pairwise d in the 0.64 to 0.85 range in secondary reports. Self-report fear did not follow.

The working summary is blunt. In the scanner and on the skin, labelling often turns a dimmer down. On the rating form, the same people may report no change, a small change, or — if the picture was only mildly unpleasant — a change in the wrong direction.

6.5 Moderators

Table 3. Pre-specified and discovered moderators

Moderator	What we asked	What we found	Poolable?
Lab vs field	Does the effect leave the laboratory?	One large observational Twitter study (Fan et al., 2019). No randomised field trial of intensity.	No
Label granularity	Does a finer word produce a larger drop?	No trial crossed a granularity score with a labelling contrast. Closest signal: more anxiety and fear words during exposure tracked larger SCR drops (Kircanski et al., 2012; Niles et al., 2015).	No
First vs second language	Do L1 labels outperform L2 labels?	Zero eligible trials, 2000–2026.	No
Stimulus intensity	Does strength of the feeling change the sign?	Yes. High intensity: downregulation. Low intensity: upregulation in Levy-Gigi Experiment 2. Lieberman 2011 intensity slices lean the same way.	Narrative only
Valence	Does labelling shrink positive feelings too?	Lieberman 2011 Study 4: labelling reduced pleasure ($t(21) = 2.21$, $n = 22$). Vlasenko 2021: labelling *increased* positive intensity. Unresolved.	No
Label type	Must the word name *my* feeling?	Objective or content labels sometimes equal or beat subjective self-labels (McRae, 2010; Constantinou, 2014; Vlasenko, 2021).	Narrative only

Moderator	What we asked	What we found	Poolable?
Outcome channel	Do body and rating move together?	Often no, in exposure paradigms.	Not a pool; a finding
Timing	Must the word arrive in the moment?	Levy-Gigi Experiment 1: simultaneous, immediate and 10-second delay did not differ.	One study

Granularity deserves a longer paragraph because it is the question everyone wants to be true.

Lisa Feldman Barrett's line of work says that people who distinguish neighbouring unpleasant states regulate better. They drink less as a hammer. They aggress less as a hammer. They show less neural reactivity to rejection (Kashdan et al., 2015). That is a between-person skill measured across days. It is not the same design as "give this person a more precise word on this trial and watch intensity fall more". Coaches collapse the two because the collapse is useful. Useful is not pooled. One pre-specified question was whether granularity moderates the labelling effect. The honest answer is that the experiment has not been done often enough to meta-analyse. An OSF preregistration on labelling and granularity exists. Usable published results were not confirmed.

Language is the other wanted moderator. India runs on more than one tongue before lunch. Matsumoto and Assar (1992) found that some Indian bilinguals recognised certain emotions better in English than in Hindi. That paper is pre-2000 and is about recognition, not regulation. It cannot be drafted into this pool. A product that offers 23 languages is making a service choice. It is not, yet, making an evidence-based language choice.

6.6 Publication bias

The restricted primary pool is too small for a confident funnel. What we can see is mild small-study lean: the tiniest cells are not the tiniest effects. Torre and Lieberman (2018) and Shukla and Mishra (2026) already warned that successful labelling studies are easier to find than quiet ones. Vlasenko et al. (2021) and Nook et al. (2021) are the counterweights that survived peer review. They pull the story toward "sometimes" and away from "always". That is the correct direction.

6.7 Risk of bias

Most laboratory studies are within-subjects, unblinded for the rating, run on campus samples, and clustered around a small set of laboratories. People who know they have just chosen the word "angry" are not blind to the independent variable. Physiology is harder to fake than a Likert box. It is not immune to demand.

RoB 2-style judgements:

- Lieberman et al. (2011), Studies 1–3: some concerns. Counterbalanced picture tasks; unblinded self-report; UCLA cluster.
- Burklund et al. (2014): some concerns. Blocked fMRI; unblinded ratings; older-adult convenience sample, mean age 65.
- Levy-Gigi and Shamay-Tsoory (2022): some concerns. MINI-screened non-clinical campus sample; pre-registered intensity and timing questions; better reporting than most.

- Tabibnia et al. (2008): some concerns to high. Small SCR *n*; some labels unrelated to the picture.
- Matejka et al. (2013): some concerns. Small SCR *n*; speech task confounds breath and word.
- Kircanski et al. (2012) and Niles et al. (2015): some concerns. Randomised applied designs; analogue anxious; self-report null, physiology better instrumented. Sensitivity only for the primary claim.
- Fan et al. (2019): not randomised. Confounding by the life events that produced the tweet.

Overall: some concerns to high for self-report. Some concerns for physiology. High for any leap from IAPS pictures to an Indian workday.

6.8 GRADE

Table 4. Certainty of evidence

Outcome	Certainty	Why
Self-report intensity, high-arousal negative, lab pictures, label vs watch	Low to moderate	Consistent in the restricted pool (g = 0.41). Downgraded for unblinded ratings, WEIRD samples, and indirectness to workplace stress.
Self-report intensity in applied or exposure settings	Low	Two analogue RCTs found no fear-rating effect.
Autonomic arousal after labelled exposure	Moderate in analogue fear; low in unselected adults	Delayed SCR reductions replicate; samples are analogue-anxious; metrics differ.
Amygdala downregulation during labelling	Moderate	Lieberman 2007 d = 0.87; Burklund 2014; Brooks 2017 imaging meta-analysis. Not the same construct as felt intensity.
Granularity as moderator of labelling → intensity	Very low	No eligible experimental test.
First vs second language of labelling	Very low	No eligible trial.
Lab vs field as pooled moderator	Very low	One observational field study.
Positive-emotion intensity	Very low	Opposite signs in Lieberman 2011 Study 4 and Vlasenko 2021.

07 Discussion

PULL QUOTE

In two applied studies the skin quietened and the fear rating did not. The body and the story can disagree.

7.1 What the number means

$g = 0.41$ is not a makeover. It is a dimmer. In machine-learning terms, you added a label and the loss dropped a bit. You did not finish training. In a Hyderabad review meeting, a manager's idea is waved away. She says "that's embarrassment" to herself, and the heat in her face drops a fraction. The comment is still in the minutes. She has put a little space between "I" and "the wave". The wave is still a wave. In start-up terms, you shipped a one-line dashboard. Revenue did not triple. The meeting got shorter.

The restricted laboratory pool says that, for strong unpleasant pictures, naming beats watching by about two-fifths of a standard deviation. The same literature says three other things that a responsible programme has to hold at the same time.

One. The body often moves when the rating does not. Kircanski's spider-fearful adults sweated less a week later if they had named the fear in the room. They did not claim to feel less afraid. Niles's speakers showed a steeper skin-conductance recovery. Their fear ratings did not separate. Matejka's speakers produced quieter skin and quieter pitch while insisting that talking about facts had helped more. Anyone who has done a hard conversation already knows this split. The hands stop shaking before the story updates. Coaches who treat the rating as the only outcome will miss half the result. Coaches who treat the rating as irrelevant will gaslight the client.

Two. Intensity flips the sign. Levy-Gigi and Shamay-Tsoory (2022) is the study that should sit on the wall of every content team. Label a high-arousal aversive picture and distress falls. Label a low-arousal aversive picture and distress rises. Lieberman et al. (2011) already had a milder version of that pattern in intensity slices. The folk advice to "name whatever you feel, whenever you feel it" is too generous. Naming a faint irritation can polish it. Naming a spike can cut it. That is closer to how exposure works than to how affirmation works.

Three. Labelling is not a master key for every other skill. Nook et al. (2021), replicated in 2026, found that naming before reappraisal made reappraisal worse. One reading is constructionist: a word crystallises a state, and a crystallised state is harder to rewrite. If that reading holds, a coaching sequence that says "first name it, then reframe it" is not obviously the right order. A sequence that says "name it and stop" may be cleaner. This paper cannot settle the order. It can stop programmes from stacking techniques as if they were Lego.

Torre and Lieberman (2018) called labelling implicit regulation because people do not believe it will work while it is working. That remains the most useful sentence in the review literature. It also explains a product problem. Users will not tap a button that feels pointless. The job of the coach, human or AI, is to tell the truth about the size of the effect and the oddness of the feeling. A small effect that people distrust is still a small effect. Selling it as alchemy is how a platform loses the next year.

7.2 Limitations

The protocol was not registered. Authorship of the source literature is clustered. Several extractable effects come from one extended laboratory family. Campus samples dominate. Pictures dominate. English dominates. Within-subjects ratings are unblinded. The primary pool is a restricted set. Its I^2 of 0% would not survive the low-intensity reversal, the failed negative replications, or the analogue self-report nulls. Physiology cannot be expressed as one honest g without pretending that an amygdala contrast and a recovery slope are the same object. Granularity and language were pre-specified and then found empty. Field evidence is one observational social-media study. India is absent.

A second limitation is conceptual. “Affect labelling” in this literature is a family, not a species. Sometimes the participant names their own feeling. Sometimes they name the face on the screen. Sometimes they pick an unrelated negative word. Sometimes they talk in full sentences next to a tarantula. Pooling those acts is already a stretch. We stretched only where the contrast was label versus watch for high-arousal negative intensity and the statistic could be checked.

A third limitation is cultural. An IAPS picture of a burn victim is not a missed promotion in Pune. It is not a family WhatsApp group at midnight. It is not a festival week with three deadlines. Indirectness is not a footnote for a product that lives on Indian phones.

7.3 What this means for an Indian workplace programme

A claims officer in Pune reads a curt email from her manager just before logging off. Before she replies, she types one word: “humiliated”. That is a legal daily rep. It costs seconds. It needs no room, no diagnosis and no 50-minute hour. For strong unpleasant affect, the laboratory best estimate is $g = 0.41$ against watching. That is worth teaching. It is not worth mythologising.

Teach it as a high-arousal tool. The use case is the team lead whose deck has just been torn apart in the quarterly review: “I am furious about this slide.” The analyst who is “slightly off this morning” may not need it. Do not follow the word straight away with “now reframe it”; one careful line of work says the word can stiffen the state you then try to rewrite. Prefer “humiliated” or “dreading Friday” to “bad”, because granularity predicts better regulation in daily life, and because more fear-words tracked more skin-conductance drop in one exposure study — not because a pooled interaction exists. It does not.

A supervisor in Chennai feels the jolt in Tamil and types it in English, because the app opened in English. Do not assume English carries the effect into Hindi, Tamil, Bengali, Marathi or any of the other languages on the product map. That trial has not been run. Running it would be the most useful next paper this platform could fund. Until then, offer the first language the person actually thinks in, and see that as a design ethic rather than as a proven moderator.

This is coaching practice. It is a countable rep. It is not therapy.

7.4 Research gaps, including the India gap

The India gap is not a rhetorical device. CTRI returned no relevant trial. ClinicalTrials.gov returned a terminated Karolinska wait-list study and a PTSD programme, both outside scope. Indian laboratories have built recognition sets, micro-expression sets, film-clip batteries and Hindi emotion lexicons. They have not, on this search, randomised a labelling practice against a control and measured intensity or arousal in working adults. A 2026 narrative review from Allahabad mapped the Western literature and again did not pool. The first Indian pool cannot be computed from zero trials.

Other gaps follow the empty moderator cells. First versus second language. Granularity as an experimental moderator, not a personality correlate. Workplace field RCTs with delayed autonomic outcomes. Positive emotion, which currently points both up and down. Dose: one word versus a 20-second utterance versus a week of daily reps. Delivery channel: screen, voice note, live group. Sex, age and caste or class context — untested here. Independent laboratories outside the original cluster.

A decent next experiment in Mumbai or Bengaluru would look like this. Working adults, not students. A real stressor, not only IAPS. Labels in the person’s first language and in English, randomised. High and low intensity arms. A granularity measure taken before the task. Self-report and skin conductance at once and at one week. A watch control and a reappraisal control. Pre-registration on CTRI and OSF. That paper would be more useful to this product than another proverb.

The last scene is shorter than the first. A coach asks a tired manager what she wants from the call. She has to say it: “I want to stop dreading Mondays.” Naming the want does not fix the Monday. It is the door the next step uses. Affect labelling, on present evidence, is a door. It is not the whole house.

08 FAQ

Q1 Does naming a feeling make it smaller?

For strong unpleasant feelings in laboratory picture tasks, yes. Six experiments, 261 adults, Hedges $g = 0.41$ against watching. For mild irritation, one experiment found the opposite. For speeches and spiders, the rating often stays put while the skin quiets.

Q2 Should I name it in English or in my first language?

We do not know. No eligible trial from 2000 to 2026 tested first versus second language. Use the language in which the feeling actually arrives. That is a practical rule, not a pooled effect.

Q3 Is this therapy?

No. It is a brief labelling practice. A rep. A coaching move. It does not diagnose, treat or replace care.

Q4 Why do I still feel the same after I name it?

Because the rating and the body disagree more often than slide decks admit. In two applied studies the sweat fell and the fear score did not. Give the nervous system a minute. Do not treat an unchanged slider as proof that nothing happened.

Q5 Does a more precise word work better?

Probably, and that has not been pooled. People with finer everyday distinctions regulate better in correlational work. In one spider study, more fear-words tracked a bigger skin-conductance drop. That is a clue, not a meta-analytic moderator.

09

References

Barrett, L. F., Gross, J., Christensen, T. C., and Benvenuto, M. (2001). Knowing what you're feeling and knowing what to do about it: Mapping the relation between emotion differentiation and emotion regulation. *Cognition and Emotion, 15*(6), 713–724. https://doi.org/10.1080/02699930143000239

Brooks, J. A., Shablack, H., Gendron, M., Satpute, A. B., Parrish, M. H., and Lindquist, K. A. (2017). The role of language in the experience and perception of emotion: A neuroimaging meta-analysis. *Social Cognitive and Affective Neuroscience, 12*(2), 169–183. https://doi.org/10.1093/scan/nsw121

Burklund, L. J., Creswell, J. D., Irwin, M. R., and Lieberman, M. D. (2014). The common and distinct neural bases of affect labeling and reappraisal in healthy adults. *Frontiers in Psychology, 5*, 221. https://doi.org/10.3389/fpsyg.2014.00221

Constantinou, E., Van Den Houte, M., Bogaerts, K., Van Diest, I., and Van den Bergh, O. (2014). Can words heal? Using affect labeling to reduce the effects of unpleasant cues on symptom reporting. *Frontiers in Psychology, 5*, 807. https://doi.org/10.3389/fpsyg.2014.00807

Fan, R., Varol, O., Varamesh, A., Barron, A., van de Leemput, I. A., Scheffer, M., and Bollen, J. (2019). The minute-scale dynamics of online emotions reveal the effects of affect labeling. *Nature Human Behaviour, 3*, 92–100. https://doi.org/10.1038/s41562-018-0490-5

Kassam, K. S., and Mendes, W. B. (2013). The effects of measuring emotion: Physiological reactions to emotional situations depend on whether someone is asking. *PLOS ONE, 8*(6), e64959. https://doi.org/10.1371/journal.pone.0064959

Kashdan, T. B., Barrett, L. F., and McKnight, P. E. (2015). Unpacking emotion differentiation: Transforming unpleasant experience by perceiving distinctions in negativity. *Current Directions in Psychological Science, 24*(1), 10–16. https://doi.org/10.1177/0963721414550708

Kircanski, K., Lieberman, M. D., and Craske, M. G. (2012). Feelings into words: Contributions of language to exposure therapy. *Psychological Science, 23*(10), 1086–1091. https://doi.org/10.1177/0956797612443830

Levy-Gigi, E., and Shamay-Tsoory, S. (2022). Affect labeling: The role of timing and intensity. *PLOS ONE, 17*(12), e0279303. https://doi.org/10.1371/journal.pone.0279303

Lieberman, M. D., Eisenberger, N. I., Crockett, M. J., Tom, S. M., Pfeifer, J. H., and Way, B. M. (2007). Putting feelings into words: Affect labeling disrupts amygdala activity in response to affective stimuli. *Psychological Science, 18*(5), 421–428. https://doi.org/10.1111/j.1467-9280.2007.01916.x

Lieberman, M. D., Inagaki, T. K., Tabibnia, G., and Crockett, M. J. (2011). Subjective responses to emotional stimuli during labeling, reappraisal, and distraction. *Emotion, 11*(3), 468–480. https://doi.org/10.1037/a0023503

Matejka, M., Kazzer, P., Seehausen, M., Bajbouj, M., Klann-Delius, G., Menninghaus, W., Jacobs, A. M., Heekeren, H. R., and Prehn, K. (2013). Talking about emotion: Prosody and skin conductance indicate emotion regulation. *Frontiers in Psychology, 4*, 260. https://doi.org/10.3389/fpsyg.2013.00260

McRae, K., Taitano, E. K., and Lane, R. D. (2010). The effects of verbal labelling on psychophysiology: Objective but not subjective emotion labelling reduces skin-conductance responses to briefly presented pictures. *Cognition and Emotion, 24*(5), 829–839. https://doi.org/10.1080/02699930902797141

Niles, A. N., Craske, M. G., Lieberman, M. D., and Hur, C. (2015). Affect labeling enhances exposure effectiveness for public speaking anxiety. *Behaviour Research and Therapy, 68*, 27–36. https://doi.org/10.1016/j.brat.2015.03.004

Nook, E. C., Satpute, A. B., and Ochsner, K. N. (2021). Emotion naming impedes both cognitive reappraisal and mindful acceptance strategies of emotion regulation. *Affective Science, 2*, 187–198. https://doi.org/10.1007/s42761-021-00036-y

Shukla, M., and Mishra, A. (2026). Effectiveness of affect labelling as an emotion regulation strategy in individuals with high emotional reactivity: A systematic review. *Current Psychology, 45*, 683. https://doi.org/10.1007/s12144-026-09229-9

Tabibnia, G., Lieberman, M. D., and Craske, M. G. (2008). The lasting effect of words on feelings: Words may facilitate exposure effects to threatening images. *Emotion, 8*(3), 307–317. https://doi.org/10.1037/1528-3542.8.3.307

Torre, J. B., and Lieberman, M. D. (2018). Putting feelings into words: Affect labeling as implicit emotion regulation. *Emotion Review, 10*(2), 116–124. https://doi.org/10.1177/1754073917742706

Vlasenko, V. V., Rogers, E. G., and Waugh, C. E. (2021). Affect labelling increases the intensity of positive emotions. *Cognition and Emotion, 35*(7), 1350–1364. https://doi.org/10.1080/02699931.2021.1959302

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 in Mumbai. He trained at Seth G.S. Medical College and holds a DPM from the College of Physicians and Surgeons, Mumbai. He leads EmotionApp at Tranquilo Meditek Sytems Private Limited: daily emotional-fitness reps on WhatsApp, 23 Indian languages, India-resident data, DPDP-compliant, ISO 9001 and ISO 13485.

HOW TO CITE

Aswani A. Name It to Tame It, Tested: Affect labelling and emotional intensity: a meta-analysis of experimental and field studies. EmotionApp Research Paper 6. Mumbai: Tranquilo Meditek Sytems Private Limited; 2026. https://emotionapp.in/research/affect-labelling-emotional-intensity-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.