International Journal of TESOL Studies (2026)
Vol. 8 (3) 134-157 https://doi.org/10.58304/ijts.260401

Article

Teachers' Productive Emotion Vocabulary: Frequency Effects and the Roles of Language Proficiency and Trait Emotional Intelligence

Allen Jie Ein Chee*
University of Nottingham Malaysia, Malaysia

Csaba Zoltan Szabo
University of Edinburgh, UK

Sharimila Ambrose
University of Nottingham Malaysia, Malaysia

Received: 2 January 2026/Received in revised form: 20 March 2026/Accepted: 2 April 2026/
Available online: 22 April 2026

Abstract
Teachers are expected to model emotional literacy and manage classroom affect (Teng, 2017), yet little is known about their ability to produce nuanced emotion vocabulary. In an online survey of 101 ESL/EFL pre- and in-service teachers, we administered a 25-item vignette-based productive emotion vocabulary task (PEVST; two responses per item) alongside LexTALE (proficiency) and the TEIQue-SF (trait EI). Teachers produced high-frequency emotion vocabulary accurately (≈78–81% in the 1k–3k bands), but performance declined for mid-/lower-frequency items (≈45% at 4k/6k, 26% at 9k), indicating strong frequency constraints. LexTALE showed a moderate positive association with PEVST accuracy (about r ≈ .41–.43, depending on the scoring method). In OLS regression (R^2 = .24), LexTALE emerged as the strongest predictor, while Emotionality contributed a smaller positive effect; the LexTALE × Emotionality interaction did not improve model fit, supporting additive rather than moderating effects. The results suggest that even largely advanced teachers may have limited productive access to lower-frequency emotion vocabulary, motivating more explicit teacher-education and curricular attention to nuanced affective labelling.

Keywords
Emotion vocabulary, language teaching, language assessment, emotion recognition, emotional intelligence

1 Introduction

Teachers work with students with an array of emotions in the classroom. Emotions shape teaching, classroom climate, and student outcomes, and teachers are expected to manage this emotional landscape

*Corresponding author. Email: Allen.Chee@nottingham.edu.my

alongside delivering content (Gkonou & Miller, 2020; Teng, 2017). Their ability to recognise, articulate, and regulate emotions significantly impacts their professional practice, classroom management, and academic success (Aldrup et al., 2024; Chen & Guo, 2020; Dewaele et al., 2018; Frenzel et al., 2021; Pugh, 2008; Richards, 2022; Song, 2018; Vince, 2016). These skills are closely intertwined with the concept of emotional intelligence (EI), which encompasses key skills such as recognising, understanding, managing and expressing emotions in oneself and others (Petrides et al., 2018; Petrides & Mavroveli, 2018). We adopt a trait EI framework, noting that trait EI is conceptually distinct from ability EI (Mayer et al., 2016) but theoretically linked to the propensity to perceive and express emotions via language (Petrides & Mavroveli, 2018). One fundamental aspect of EI is emotion vocabulary, specifically, the knowledge and use of emotion words to label and express emotional experiences. As language of emotions shapes how individuals perceive and process their emotional experiences, limited emotion vocabulary can constrain teachers' capacity to comprehend classroom emotional dynamics and to model precise, context-appropriate expression (Barrett, 2017).

From a language learning perspective, expressing different degrees and nuances of emotion presupposes access to a sufficiently broad and precise emotion lexicon. Studies have shown that learners with limited emotion vocabulary tend to rely on general or superordinate terms (e.g., "happy") rather than specific hyponyms that capture finer distinctions (e.g., "thrilled"), thereby restricting their ability to convey emotional nuances accurately (Bazhydai et al., 2019; Ebert et al., 2014). An insufficient emphasis on emotion vocabulary in English as a Foreign Language (EFL) classrooms and teaching materials has been suggested as one reason for underdeveloped emotional expression skills in the target language (Sánchez & Pérez-García, 2020). While definitions and synonyms can be learned from dictionaries and thesauruses, there is no guarantee that those emotion words can be used appropriately within situated contexts (Liu & Zhong, 2016). Given that teacher input is central to language learning (Gaies, 1983; Gass & Madden, 1985; Long, 2020; Wang, 2022), teachers themselves play a crucial role in shaping learners' emotion vocabulary and broader EI (Frenzel et al., 2021; Schonert-Reichl et al., 2017).

Despite the importance of possessing a good mastery of emotion vocabulary knowledge, limited studies have specifically assessed teachers' emotion vocabulary knowledge. Most research focuses on EI or emotion regulation and their effects on burnout, teaching pedagogies and student outcomes (see Alam & Ahmad, 2018; Anari, 2012; Arens & Morin, 2016; Curci et al., 2014; Dewaele et al., 2018; Lee et al., 2019; Mérida-López & Extremera, 2017; Molero et al., 2019; Pourbahram & Hajizadeh, 2018; Thomas et al., 2012; Valente et al., 2020; Wang, 2022), without considering the precise language teachers use to express emotions. While these studies provide valuable insights into teachers' EI, they do not directly assess the linguistic resources teachers draw upon to perceive, differentiate, and express emotions. In particular, there remains a lack of research that directly measures teachers' productive emotion vocabulary knowledge, profiles the frequency bands of emotion words teachers are able to retrieve, and examines whether individual differences such as language proficiency and trait EI account for variance in emotion vocabulary performance. As a result, little is known about the breadth, depth, and accessibility of the specific emotion lexicon available to teachers, or the factors that may shape its development.

Recent work on the measurement of emotion vocabulary among English L1/L2 speakers has begun to address this gap by developing validated tools to assess productive emotion-word labelling using contextualised vignettes (Chee et al., 2025; Szabo et al., forthcoming). Findings from this research suggest that even individuals with upper-intermediate to advanced English proficiency tend to rely on high-frequency, general emotion words (e.g., "sad", "happy", "angry") rather than lower-frequency, more precise alternatives, revealing a general difficulty in accessing nuanced emotion vocabulary. Such evidence highlights that emotion vocabulary knowledge cannot be assumed even among proficient speakers. However, it remains unknown whether teachers, whose emotional language plays a critical pedagogical role, possess sufficient breadth and depth of emotion vocabulary to accurately and effectively express emotions in classroom contexts. While previous work supported lexical availability assessment for pre-service English teachers (Quintanilla & Kloss Medina, 2024) and validated the

PEVST and documented productive emotion vocabulary labelling among general English L1/L2 users (see Chee et al., 2025; Szabo et al., forthcoming), teachers represent a distinct population whose emotion language is pedagogically consequential as their role requires frequent, pedagogically-shaped affective talk (e.g., feedback, classroom management, modelling emotional literacy). Specifically, teachers are expected to model emotional literacy, manage classroom emotional climate, and deliver feedback using emotion labels. However, their productive emotion vocabulary knowledge has rarely been assessed directly. Therefore, this study investigates teachers' productive emotion-word labelling in contextualised vignettes and examines whether language proficiency and EI uniquely account for variance in productive emotion vocabulary.

2 Literature Review

2.1 Global policies and curriculum on emotional intelligence and emotion vocabulary

Curricular and policy frameworks worldwide increasingly emphasise teachers' roles in fostering students' social–emotional competencies, including the precise labelling and expression of emotions. For instance, the Common European Framework of Reference for Languages (CEFR) specifies that learners at the B2 level should be able to express a broad range of emotions (Council of Europe, 2020), reflecting a clear expectation that emotional literacy is part of communicative competence. If curricula expect learners at B2 to express a broad range of emotions, teachers also need sufficient affective lexical resources to model and scaffold this discourse. Similarly, although not language curricula, social and emotional frameworks such as Promoting Alternative Thinking Strategies (PATHS; Humphrey et al., 2015) and Recognising, Understanding, Labelling, Expressing and Regulating (RULER; Brackett & Cipriano, 2020) explicitly teach the recognition, labelling, and regulation of emotions as core skills necessary for students' personal and academic development; with emotion vocabulary being a core tenet.

National education policies also reflect this growing emphasis on EI and emotional development. In Australia, the New South Wales Department of Education mandates that all pre-service teachers undergo psychometric testing to ensure they demonstrate sufficient levels of EI before entering the profession (New South Wales Government, 2018). This initiative ensures that teachers are equipped with the necessary skills and competence to effectively manage the emotional complexities of teaching (Brown, 2014; Frenzel et al., 2021; Teng, 2017). Similarly, the Malaysian Education Blueprint for 2013-2025 also articulates a commitment to developing "the child holistically along intellectual, spiritual, emotional, and physical dimensions..." (Ministry of Education Malaysia, 2013, p. E-4). The English Language Management Guidebook (Buku Panduan Bahasa Inggeris) further supports this goal by advocating for students to be assessed holistically (cognitive, affective, and psychomotor) (Ministry of Education Malaysia, 2019).

International and Malaysian policies collectively signal a broad consensus: emotional development and emotional literacy; particularly the ability to recognise, understand, and express emotions through language are essential components of modern education. Yet, despite these policy ambitions, teachers often receive little training in how to operationalise such goals through instruction, often limited to suggestions such as providing "statements to calm pupils' emotions and prepare them for the next lesson" at the end of a lesson (Ministry of Education Malaysia, 2019, p. 43). While policy frameworks expect emotionally intelligent teaching and outcomes linked to emotion vocabulary, pre-service teacher education rarely addresses these linguistic and emotional competencies in practical terms (Corcoran & O'Flaherty, 2022; Schonert-Reichl, 2017; Turner & Stough, 2020).

Operationalising 'emotional literacy' in classrooms arguably depends on teachers' capacity to recognise and label affective states with lexically precise and context-appropriate language. Such capacity may be shaped by at least two interrelated factors. First, language proficiency constrains the

breadth, depth, and accessibility of lexical resources available for affect labelling, particularly in contexts that require nuance and clarity. Second, trait EI (especially dimensions pertaining to emotion perception and expression) may influence teachers' sensitivity to affective cues and their ability to identify nuanced emotions that may require specific lexical encoding. Taken together, these factors provide a theoretically grounded basis for examining productive affect-related vocabulary as an outcome shaped by both linguistic and emotional capacities.Yet, the linguistic dimension of EI, productive affect-related vocabulary, has received comparatively little empirical attention in teacher populations. As teachers are uniquely positioned as classroom language models and emotional socialising agents (Frenzel et al., 2021; Schonert-Reichl et al., 2017), limitations in their productive emotion vocabulary may affect both instructional discourse (e.g., feedback framing) and the modelling of emotional literacy. Therefore, this study addresses that gap by examining teachers' performance on a productive affect-related vocabulary task (PEVST) and its associations with language proficiency and trait EI.

2.2 Implications of teachers' use of emotion vocabulary

Teaching is widely regarded as a profession that involves substantial emotional labour (Gkonou et al., 2020; Gkonou & Miller, 2019; Hargreaves, 2001; Isenbarger & Zembylas, 2006; King, 2016; Loh & Liew, 2016; Nyanjom & Naylor, 2021; Yoo & Carter, 2017). Teachers are expected to adhere to implicit emotional display rules, which require them to exhibit positive emotions, often at the expense of their well-being (Chang, 2020; Frenzel et al., 2021; Taxer & Frenzel, 2015). However, studies in educational psychology and emotion research increasingly recognise that negative emotions should not be suppressed; rather, they can serve as pedagogical resources when articulated through precise and context-appropriate language (Song, 2018; Zembylas, 2007). In this regard, a teacher's emotion vocabulary becomes a crucial linguistic tool for emotional regulation and classroom communication. For example, when a teacher expresses anger in an unregulated manner (e.g., 'I'm so angry with you!), the lexical choices convey a negative attribution toward the student's character, potentially damaging motivation and self-esteem. In contrast, a more mindful use of emotion vocabulary (e.g., "I'm disappointed about this outcome. Let's work on improving this together."), expressing empathy and sympathy, conveys the belief that the student's struggle stems from temporary external challenges – such as lack of skills or knowledge – that can be improved with support (Frenzel et al., 2021). This mindful shift in emotion vocabulary use reframes the situation from personal failure to problem solving, guiding students to interpret emotional feedback as constructive rather than punitive (Aldrup et al., 2024). The effectiveness of such exchanges thus depends not just on teachers' emotional regulation, but also their ability to select emotion words that align intent with pedagogical goals.

Teachers' affective lexical choices (productive emotion vocabulary) shape how students interpret feedback and model emotion discourse making productive access to precise emotional labels pedagogically consequential. Context-appropriate and precise use of emotion vocabulary supports constructive feedback (e.g., disappointment versus anger), while overreliance on generic terms (e.g., sad, angry, happy) may potentially obscure intent or cause miscommunication. As primary socialising agents in the classroom, teachers model how emotions are named, differentiated, and managed (Valiente et al., 2020). However, doing so effectively requires access to a wide and nuanced knowledge of emotion vocabulary, often lower frequency lexical items that are harder to retrieve (Chee et al., 2025; Szabo et al., forthcoming). From a psycholinguistic perspective, this tendency to default to more common emotion labels may also reflect Zipf's Principle of Least Effort, which posits that individuals are more likely to select linguistic forms that require the least cognitive effort to retrieve and produce (Zhu et al., 2018; Zipf, 2016). Consequently, unless teachers possess sufficiently developed affective lexical resources, they may rely disproportionately on higher frequency emotion terms, limiting the precision with which emotional states are communicated.

Evidence from recent emotion vocabulary studies underscores this challenge. Using the Productive Emotion Vocabulary Size Test (PEVST), Chee et al. (2025) found that even participants with upper-intermediate to advanced English proficiency struggled to produce mid- and low-frequency emotion words, despite performing well on general proficiency measures such as LexTALE (Lemhöfer & Broersma, 2012). Participants tended to rely on high-frequency, generic emotion terms (e.g., "hungry" instead of "famished"), revealing limited productive access to nuanced vocabulary. The PEVST, designed to assess productive emotion vocabulary across different frequency bands, provides a validated means of distinguishing general lexical proficiency from emotion-specific lexical competence. Emotion vocabulary is defined according to Pavlenko's (2008) classification of emotion-, emotion related-, and emotion-laden words. If proficient language users display such limitations, it remains uncertain whether teachers, who are expected to model and scaffold the use of emotion vocabulary, possess sufficient breadth and depth of emotion vocabulary to meet these demands.

Thus, mastery of emotion vocabulary underpins teachers' ability to align emotional expression with pedagogical intent, sustain emotionally supportive environments, and model adaptive regulation for students (Bielak & Mystkowska-Wiertelak, 2022; Cheng et al., 2022; Gallingane & Han, 2015; Jennings & Greenberg, 2009; Schonert-Reichl et al., 2017; Wang, 2022). Despite its importance, little is known about teachers' ability to recognise and produce emotion vocabulary, or how this lexical competence interacts with broader constructs such as EI and language proficiency. From the trait EI perspective, the Emotionality factor (comprising emotion perception and expression) is the facet most theoretically proximal to the productive labelling of affective states (Petrides, 2009; Petrides et al., 2018; Petrides & Mavroveli, 2018). Importantly, the demands of the productive emotion vocabulary size task (PEVST) align closely with these factors. Participants must first infer the emotional state implied in contextualised vignettes, a process that draws on emotion perception and sensitivity to affective cues, and subsequently produce an appropriate lexical label for that state, which reflects emotion expression in linguistic form. In contrast, other trait EI factors within the TEIQue framework, such as Sociability or Self-control, primarily relate to interpersonal influence, assertiveness, or regulation of emotional reactions, and are therefore may have lesser implications in the lexical identification and labelling of emotions required by the task.

2.3 Research questions

Given policy expectations and the pedagogical role of affective talk, it is important to establish teachers' baseline productive affective lexicon and its constraints by frequency. Building on PEVST work, we test whether language proficiency and trait Emotionality predict performance of teachers' productive emotion-word labelling through contextualised vignettes. This study aims to examine ESL/EFL teachers' ability to recognise and produce emotion vocabulary across different frequency bands and to explore how emotional intelligence and language proficiency predict this knowledge. Accordingly, this study is guided by the following research questions:

1. To what extent do ESL/EFL teachers demonstrate productive emotion vocabulary knowledge across different frequency bands?
2. To what extent do language proficiency and trait EI explain individual differences in emotion vocabulary production?

3 Methodology

3.1 Sampling

Snowball sampling was used to recruit participants to target a more general population of teachers. All data were collected via an online survey. Participants were teachers working in English-medium

classrooms or educational contexts where English was used as the primary language of instruction, regardless of whether English was their L1/2. A total of 183 ESL/EFL pre- and in-service teachers were recruited for the study. However, 82 teachers failed to complete all sections and were omitted from the analysis, resulting in a total of 101 teachers in this study. Table 1 shows a concise overview of the demographic characteristics of the participants in this study. This study was approved by the Faculty of Arts and Social Sciences Research Committee at University of Nottingham Malaysia [FASS2022-0016/SoEd/ACJ18823555(Revision2)]

Table 1
Summary of Participants' Demographic Characteristics

Demographics	***N***
Gender	
Male	17
Female	84
Language Background	
L1 English	35
L2 English	66
Nationality	
Malaysian	80
Non-Malaysian	21
Teacher Status	
Pre-service	32
In-service	69
Mean Age (years)	29

Note. L1 English are participants with English as their first language; L2 English are participant with English as an additional language; Non-Malaysian participants include participants from 16 countries including Vietnam, China, Korea, USA etc.

3.2 Instruments

3.2.1 Productive Emotion Vocabulary Size Test (PEVST; Appendix A)

The PEVST is a vignette-based productive task designed to elicit emotion word production. Each item consists of a short scenario describing an emotional situation (e.g., interpersonal conflict, achievement, disappointment), followed by two blank spaces. Participants were instructed to write two emotion words that best described how the person in the vignette would feel. The task was administered in written form and was untimed. The version used in this study contained 25 items selected from Chee et al. (2025) and Szabo et al. (forthcoming). Items were selected to represent both core and complex emotions and to span a range of lexical frequency. Validated items appeared in random order without cross-item restrictions (i.e., participants not advised on whether they can provide the same word across items). Prior to the main analyses, Rasch analyses were conducted to confirm that the retained items demonstrated acceptable psychometric quality in the present teacher sample (see supplementary Tables S1–S4 and Figures S1–S2 for full Rasch output). Figure 1 shows a sample item from the PEVST.

Figure 1
Sample Item from the PEVST: "hungry/famished"

"I haven't eaten all day and now my stomach is hurting. I really need to eat something!"

Write 2 words that you think the main character might **feel**.

1
2

3.2.2 Lexical Test for Advanced Learners of English (LexTALE)

LexTALE (Lemhöfer & Broersma, 2012) was used to control for participants language proficiency. Participants were required to respond to 60 lexical items, consisting of 40 words and 20 non-words, with "Yes" or "No" to determine if the word presented is a word or non-word. The score for LexTALE is calculated using the average percentage of correct responses (average % correct), which is corrected for the unequal proportion of words and nonwords in the test (Lemhöfer & Broersma, 2012). In this study, it yielded an acceptable reliability coefficient with Cronbach's alpha of 0.88. Mean LexTALE performance was 83.45% (*SD* = 12.04), indicating that most participants were within the advanced proficiency range. Table 2 shows a summary of participants' language proficiency.

Table 2
Summary of Participants' Language Proficiency as Measured by LexTALE

	Range	***N***	**Min**	**Max**	**Mean**	**SD**
Advanced	80-100	66	80.00	100.00	90.93	5.67
Upper intermediate	60-79	32	60.00	78.75	70.86	5.20
Lower intermediate	<60	3	48.75	56.25	48.38	3.28

Note. Means are used for descriptive purposes only. Exact scores as continuous variables are used in subsequent analysis.

3.2.3 Trait Emotional Intelligence Questionnaire – Short Form (TEIQue-SF)

TEIQue-SF was used in this study as a measure of EI (see Mayer et al., 2016; Takšić & Mohorić, 2008). TEIQue-SF is a 30-item questionnaire that assesses 15 specific facets (adaptability, assertiveness, emotion appraisal, emotion control, emotion expression, emotion management, low impulsiveness, relationships, self-esteem, self-motivation, social awareness, stress management, trait empathy, trait happiness and trait optimism) and four factors of EI (well-being, self-control, emotionality, and

sociability) from the original TEIQue (Petrides, 2009). Participants indicate their agreement on each item on a Likert scale ranging from 1 (completely disagree) to 7 (completely agree). Factor scores were computed following the TEIQue-SF. scoring key by averaging the items corresponding to each factor. The following items have been originally reversed scored: 2, 4, 5, 7, 8, 10, 12, 13, 14, 16, 18, 22, 25, 26, and 28. Table 3 shows a summary of descriptive statistics of teachers' trait EI score and their respective factors from the TEIQue-SF.

Table 3
Descriptive Statistics of Trait EI and Their Respective Factors from the TEIQue-SF

	N	Min	Max	Mean	SD
Trait EI	101	2.6	6.6	4.9	0.8
Well-being	101	2.5	7.0	5.4	1.0
Self-Control	101	2.3	6.3	4.4	0.9
Emotionality	101	2.9	6.9	5.1	0.9
Sociability	101	2.2	6.8	4.5	0.9

In this study, it yielded an acceptable reliability coefficient with Cronbach's alpha of 0.84 for the global score. All the other EI factors – wellbeing ($\alpha = 0.88$), self-control ($\alpha = 0.91$), emotionality ($\alpha = 0.90$), and sociability ($\alpha = 0.91$) – also yielded a good range of reliability ratings (Bonett & Wright, 2015; Taber, 2018). Out of the four factors of EI measured, the participants scored the highest in well-being and emotionality, with means of 5.4 and 5.1 respectively.

3.3 Data analysis

3.3.1 Scoring the PEVST

While the test uses a written format, the constructs tested within this test are emotion recognition and production. Obvious spelling errors were corrected for scoring purposes, while responses were otherwise left unchanged. Scoring procedures follow the steps outlined in Chee et al. (2025), using both dominant and accuracy scoring. In dominant scoring, points were awarded based on the top three most frequent responses among participants, while non-dominant responses were scored as 0. All inflected forms were grouped under a single headword. On the other hand, accuracy scoring evaluated the actual responses provided by participants. The first two authors referenced definitions using two online dictionaries, The Britannica Dictionary (Encyclopædia Britannica, 2024) and Merriam-Webster Dictionary (Merriam-Webster, 2024). Responses were coded as correct if it included synonyms of the target emotion or other acceptable emotion words that described the vignette and appeared in the emotion word list (see Chee et al., 2025), while non-obvious spelling errors leading to a different word, irrelevant responses, or words absent from the list were deemed incorrect. Each vignette elicited two responses: participants were asked to provide two emotion words that best described the scenario. Responses were scored at the word level. Each accurate emotion word was awarded 1 point, yielding a maximum possible Accuracy score of 50 (25 items × 2 responses).

3.3.2 Correlation analysis

Since vocabulary knowledge and frequency of word occurrence are interrelated (see Schmitt & Schmitt, 2020) and emotion vocabulary production has also been shown to be influenced by language

proficiency, the modulated scores enable us to better differentiate between participants (see Szabo et al., forthcoming). Specifically, three indices were derived. FREQ represents the sum of the inversed frequency values (Zipf scale) of produced emotion words based on the SUBTLEX-UK corpus (van Heuven et al., 2014), such that lower-frequency words receive higher weights. ITEMDIFF reflects the sum of inversed item difficulty scores, giving greater weight to responses produced for more difficult vignettes. These two components were then combined (FREQ × DIFF), calculated as the product of FREQ and ITEMDIFF. In this way, responses that involved both lower-frequency emotion vocabulary and more difficult items contributed more strongly to the overall score. The resulting modulated scores therefore capture not only the accuracy of emotion-word production, but also the lexical sophistication and contextual difficulty of the responses. In essence, participants with a higher language proficiency should be able to provide more precise and lower frequency emotion words. These weighted scores were used for an exploratory correlation analysis to investigate the relationship between EI, LexTALE and teachers' emotion vocabulary knowledge.

3.3.3 Regression analysis

To examine whether teachers' language proficiency and emotional intelligence independently or jointly predict productive emotion vocabulary, we modelled teachers' PEVST Accuracy (sum-correct across 25 items; range 0–50) as the outcome, with language proficiency (LexTALE) and the TEIQue-SF Emotionality factor as the key predictors, and Age, Gender, and Teacher Status (pre-/in-service) entered as covariates. Continuous predictors (LexTALE, Emotionality, Age) were z-standardised (M = 0, SD = 1) to aid interpretability and to mitigate collinearity in potential interaction models; coefficients for continuous terms can thus be read as the expected change in Accuracy associated with a one–standard deviation increase in the predictor. Models were estimated using ordinary least squares (OLS) with two-tailed α = .05 and 95% confidence intervals. Assumptions were examined with the performance package: linearity and functional form via residuals-versus-fitted and residuals-versus-predictor plots with LOESS overlays (check_model); multicollinearity via check_collinearity (generalised VIFs; VIF < 5 deemed acceptable); and homoscedasticity via check_heteroscedasticity with corroborating visual inspection of residuals-versus-fitted plots. Model performance indices (R^2, adjusted R^2, RMSE, sigma, AIC/AICc/BIC) and ranked model comparisons were obtained using compare_performance.

Data was analysed via R Studio v4.3.3 (RStudio Team, 2020) using apaTables (Stanley, 2021), psych (Revelle, 2023), performance (Lüdecke et al., 2021) and Tidyverse (Wickham et al., 2019) packages.

4 Results

4.1 Naming patterns

In total, the PEVST elicited 5,050 responses. Using dominance scoring, a total of 2,537 (50%) responses were marked as correct and matched the target emotion word or a close synonym (Table 4). 2,513 non-dominant responses were marked as incorrect.

All 25 items directly matched their target emotion word based on the top three most dominant responses. Both low frequency target items ("disbelief" and "inquisitive") had a high frequency synonym counterpart as their dominant responses. Some items have similar naming patterns in this study that matches with the items in Chee et al. (2025) and Szabo et al., (forthcoming) (e.g., "inquisitive": "curious", "excite", and "inquisitive"; "dizzy": "dizzy", "tire", and "sick"), indicating consistency in response tendencies across studies and supporting the stability of item interpretation in the revised version of the PEVST. Additionally, this also suggests that synonym preference patterns remain relatively consistent across different participant groups in different studies.

Table 4
Summary of Items That Matched Their Target Emotion Word

Target emotion	Freq	Dom 1	Count 1	NA1	Dom 2	Count 2	NA2	Dom 3	Count 3	NA3
happy	1k	happy	51	50.50	excite	21	20.79	grateful	20	19.80
regret	3k	regret	73	72.28	guilty	18	17.82	remorse	15	14.85
embarrass	2k	embarrass	59	58.42	shy	25	24.75	anxious	16	15.84
disgust	2k	disgust	80	79.21	nausea	10	9.90	annoy	7	6.93
lone	2k	lone	70	69.31	sad	43	42.57	isolate	17	16.83
worry	1k	worry	76	75.25	anxious	36	35.64	concern	13	12.87
safe	1k	safe	61	60.40	secure	20	19.80	love	17	16.83
jealous	5k	jealous	64	63.37	envious	23	22.77	envy	19	18.81
surprise	1k	surprise	68	67.33	happy	49	48.51	excite	11	10.89
boring	1k	boring	77	76.24	tire	28	27.72	exhaust	10	9.90
excite	1k	excite	77	76.24	happy	15	14.85	nervous	12	11.88
lazy	1k	lazy	67	66.34	relax	26	25.74	procrastinate	13	12.87
obsess	4k	excite	35	34.65	happy	17	16.83	obsess	9	8.91
dizzy	5k	dizzy	55	54.46	tire	30	29.70	sick	19	18.81
relieve	3k	relieve	59	58.42	relief	20	19.80	relax	18	17.82
nostalgia	6k	nostalgic	35	34.65	happy	17	16.83	reminisce	15	14.85
shy	1k	shy	49	48.51	curious	39	38.61	nervous	18	17.82
greed	4k	happy	33	32.67	satisfy	30	29.70	greed	19	18.81
disbelief	12k	doubt	26	25.74	disbelief	19	18.81	sceptic	16	15.84
hungry	1k	hungry	69	68.32	starve	28	27.72	famished	12	11.88
curious	2k	curious	87	86.14	excite	10	9.90	determine	9	8.91
impulse	4k	happy	15	14.85	desire	12	11.88	impulse	11	10.89
grateful	3k	grateful	44	43.56	love	39	38.61	happy	23	22.77
flirt	6k	flirt	24	23.76	happy	22	21.78	nervous	12	11.88
inquisitive	9k	curious	90	89.11	excite	28	27.72	inquisitive	16	15.84

Note. Freq: Frequency bands (Nation, 2004); Count 1,2 and 3: Frequency of response occurrence; Dom 1: Most dominant response; Dom 2: Second most dominant response; Dom 3: Third most dominant response; NA 1, 2, and 3: Naming agreement for Dom 1, 2, and 3 respectively.

Table 5 shows a summary of accuracy scoring for all 25 items.

Using accuracy scoring, a total of 3,495 responses (69%) were marked as correct and matched the target emotion word (Table 5), while 1,555 responses were marked as incorrect. The mean Accuracy score was 35.88 (*SD* = 4.48). This accuracy rate indicates that teachers were largely successful in identifying and producing target emotion words, however, a proportion of inaccurate responses reveals persistent lexical gaps. Such gaps may be attributed to word-frequency effects or incomplete conceptual differentiation of emotion vocabulary, supporting prior findings that even highly proficient English users tend to rely on high-frequency or superordinate emotion vocabulary (Chee et al., 2025; Szabo et al., forthcoming)

Table 5
Summary of Accuracy Scoring for All 25 Items

No	Target emotion	Freq	Examples	Sum of accurate responses
1	happy	1k	Joy, delighted, elated	174
2	regret	3k	Remorse, guilty, sorry	182
3	embarrassed	2k	Ashamed, humiliated, shameful	158
4	disgusted	2k	Repulsed, sick, gross	159
5	lonely	2k	Neglected, abandoned, rejected	181
6	worried	1k	Scared, nervous, panic	185
7	safe	1k	Comfortable, calm, protected	195
8	jealous	5k	Jealous, envious, envy	179
9	surprised	1k	Delighted, touched, shocked	194
10	boring	1k	Boring, tired, exhausted	168
11	excited	1k	Thrilled, eager, impatient	154
12	lazy	1k	Sluggish, demotivated, tired	127
13	obsessed	4k	Enthusiastic, determined, mesmerised	81
14	dizzy	5k	Weak, unwell, nauseous	131
15	relieved	3k	Grateful, thankful, reassured	131
16	nostalgic	6k	Sentimental, longing, melancholic	106
17	shy	1k	Curious, confused, timid	153
18	greedy	4k	Craving, addicted, voracious	109
19	disbelief	12k	Dubious, dissatisfied, apprehensive	93
20	hungry	1k	Ravenous, unfulfilled, desperate	115
21	curious	2k	Inquisitive, intrigued, perplexed	133
22	impulsive	4k	Determined, greedy, committed	82
23	grateful	3k	Appreciative, thankful, touched	177
24	flirtatious	6k	Excited, confident, attracted	75
25	inquisitive	9k	Puzzled, fascinated, enthusiastic	53

Table 6
PEVST Accuracy Rates Across Frequency Bands

Frequency band	Items (n)	Accurate responses	Total possible responses	Accuracy rate (%)
1k	9	1,465	1,818	80.58
2k	4	631	808	78.09
3k	3	490	606	80.86
4k	3	272	606	44.88
5k	2	310	404	76.73
6k	2	181	404	44.80
9k	1	53	202	26.24
12k	1	93	202	46.04
Total	**25**	**3,495**	**5,050**	**69.21**

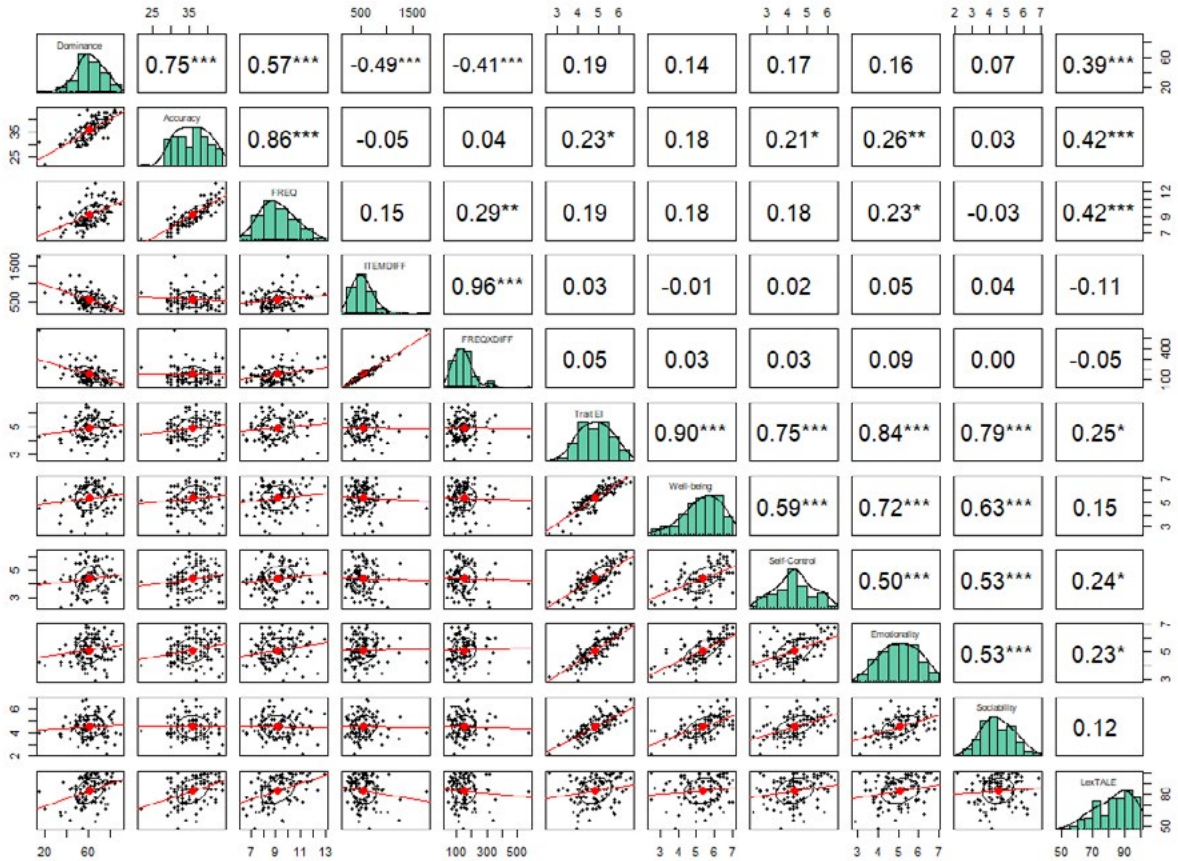
Note. Frequency bands are based on Nation's (2004) framework. Each of the 25 vignettes elicited two emotion-word responses. Therefore, the total possible responses per item = 202 (101 participants × 2 responses), and total possible responses across all items = 5,050. Accuracy rate represents the percentage of responses coded as acceptable emotion words under accuracy scoring.

As shown in Table 6, teachers performed strongly on high-frequency emotion vocabulary in the 1k–3k bands (78.09%–80.86%). In contrast, accuracy declined substantially for mid- and lower-frequency emotion vocabulary, particularly in the 4k (44.88%) and 6k (44.80%) bands. The lowest performance occurred for the 9k band item (26.24%). However, accuracy decline was not strictly monotonic across frequency bands, as the 5k band remained relatively high (76.73%). This deviation may reflect some item-level variation associated with the small number of vignettes representing certain bands, suggesting that factors beyond frequency alone that may have influenced response accuracy. Overall, these results suggest that teachers' productive access to emotion vocabulary is constrained by lexical frequency, with teachers relying more successfully on high frequency affective labels and showing reduced retrieval of less frequent, more nuanced emotion words.

4.2 Relationship between emotion vocabulary, emotional intelligence, and language proficiency

Spearman rank correlation was calculated to investigate the relationship between emotion vocabulary scores (dominant and accuracy), EI and the respective factors, and language proficiency. Figure 2 shows a summary of the correlation between emotion vocabulary, EI and LexTALE

Figure 2
Correlation Matrix between Emotion Vocabulary, EI, and Language Proficiency

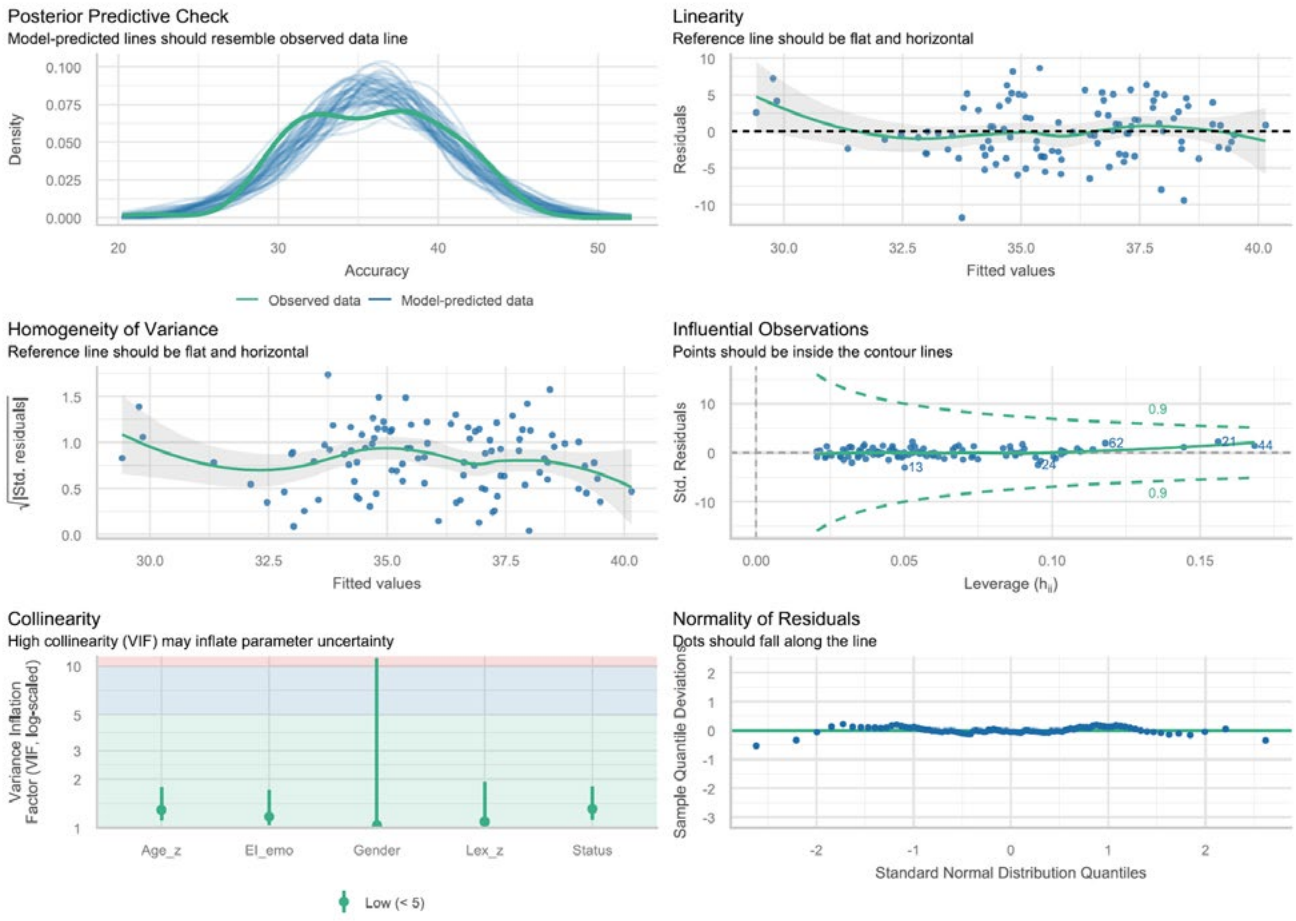


Note. Dominance = Sum of dominance scores; Accuracy = Sum of accuracy scores; FREQ = Sum of inversed frequency scores (Zipf) from SUBTLEX-UK (van Heuven et al., 2014); ITEMDIFF = Sum of inversed item difficulty; FREQxDIFF = FREQ * ITEMDIFF; LexTALE = Language proficiency scores. 95% confidence intervals appear in brackets below the correlations * $p < .05$, ** $p < .01$, *** $p < .001$

When scoring the PEVST, dominance ($r = 0.39$, $p < .001$), accuracy ($r = 0.42$, $p < .001$) and modulated accuracy scores with frequency data (Zipf) from SUBTLEX-UK (van Heuven et al., 2014) showed moderate positive correlation with language proficiency ($r = 0.42$, $p < .001$). Weighted item difficulty, however, showed no significant correlation with language proficiency. Global Trait EI ($r = 0.23$, $p < .05$) and self-control dimension ($r = 0.26$, $p < .01$) showed a weak positive correlation with the accuracy scoring. The Emotionality dimension ($r = 0.26$, $p < .01$) showed a weak positive correlation with accuracy scoring and modulated frequency scores ($r = 0.23$, $p < .05$). Given that the Emotionality factor encompasses emotion perception and expression, both critical components for identifying and producing emotion words, it was selected for the regression model in place of global Trait EI. This focus allows a more direct test of the affective processes most relevant to productive emotion vocabulary use.

Assumption checks for the regression indicated approximately linear relationships, no problematic multicollinearity (all VIFs < 2), and no evidence of heteroscedasticity (see Figure 3 below).

Figure 3
Assumption Checks for Regression Analysis



The additive model (Accuracy ~ LexTALE + Emotionality + Age + Gender + Status) explained a significant and moderate proportion of the variance in emotion vocabulary accuracy, $R^2 = .24$, adjusted $R^2 = .20$, $F(5, 95) = 6.03$, $p < .001$. Table 7 shows the detailed regression coefficients.

Table 7
Multiple Regression Predicting Teachers' Productive Emotion Vocabulary

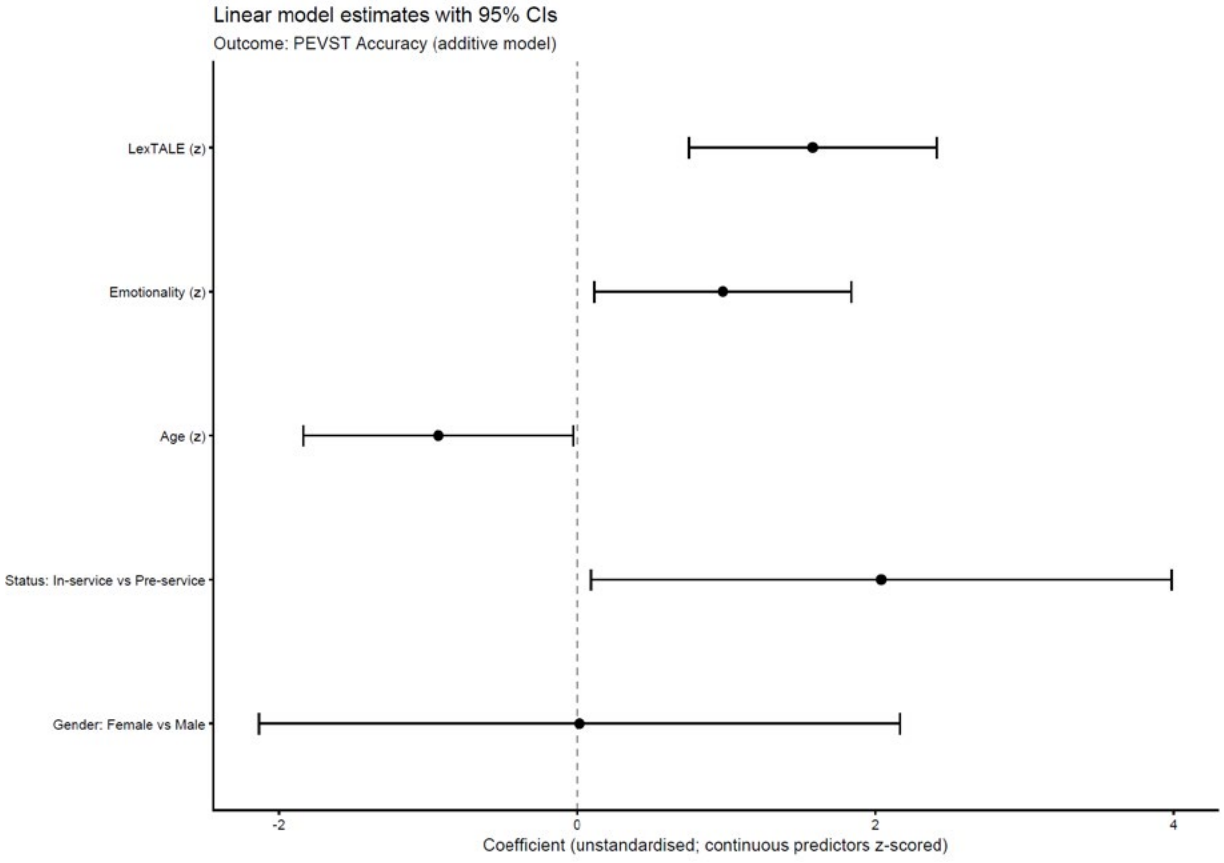
Parameter	***b***	**95% CI**	***t***	**df**	***p***	***β* (Std.)**	**Std. *β* 95% CI**
(Intercept)	34.47	[32.23, 36.72]	30.49	95	< .001***	–	–
LexTALE (z)	1.58	[0.75, 2.41]	3.78	95	< .001***	0.35	[0.17, 0.54]
EI Emotionality (z)	0.98	[0.12, 1.84]	2.26	95	.026*	0.22	[0.03, 0.41]
Age (z)	-0.93	[-1.83, -0.03]	-2.04	95	.044*	-0.21	[-0.41, -0.01]
Gender (Female)	0.02	[-2.13, 2.16]	0.01	95	.988	0.00	[-0.48, 0.48]
Status (In-service)	2.04	[0.09, 3.99]	2.08	95	.040*	0.46	[0.02, 0.89]

Model fit: AIC = 574.68; AICc = 575.88; BIC = 592.98; R^2 = .24; Adjusted R^2 = .20; σ = 4.00

Note. b = unstandardized coefficient; β = standardized coefficient.

* p < .05, ** p < .01, *** p < .001

Figure 4
Linear Model Estimates with 95% Confidence Intervals for Predictors of PEVST Accuracy (additive model)



Controlling for age and demographic factors, LexTALE emerged as the strongest positive predictor (*b* = 1.58, SE = 0.42, 95% CI [0.75, 2.41], *t* (95) = 3.78, *p* < .001; β = .35), indicating that teachers with higher proficiency produced more accurate emotion words. Emotionality also showed a positive association with Accuracy (*b* = 0.98, SE = 0.43, 95% CI [0.12, 1.84], *t* (95) = 2.26, *p* = .026; *β* = .22), suggesting that teachers with stronger emotion perception and expression performed better on the PEVST. Among covariates, Age was negatively associated with Accuracy (*b* = –0.93, SE = 0.45, *p* = .044; *β* = –.21), while Status showed a positive effect (*b* = 2.04, SE = 0.98, *p* = .040; *β* = .46), with in-

service teachers outperforming pre-service teachers. Gender had no significant effect. To test whether Emotionality moderated the proficiency–accuracy link, a LexTALE × Emotionality interaction term was added to the model. The interaction was not significant (F (1, 94) = 0.00, p = .98; $\Delta R^2 < .001$) and did not improve model fit. Simple-slopes analyses confirmed near-identical LexTALE slopes at Emotionality = −1, 0, +1 *SD* (Figure 4). These results indicate that Emotionality and proficiency exert additive rather than interactive effects on productive emotion vocabulary. In sum, language proficiency (LexTALE) shows the largest unique association with productive emotion vocabulary, with smaller positive contributions from Emotionality and teacher status, and a small negative association with age. No evidence of EI-by-proficiency moderation emerged.

5 Discussion

This study examined ESL/EFL teachers' productive emotion-related vocabulary using a revised PEVST and tested whether language proficiency (LexTALE) and trait Emotionality (TEIQue-SF) explain individual differences in performance. Overall, the findings indicate that teachers were generally successful in identifying and producing emotion vocabulary, although their performance was constrained by lexical frequency, with higher accuracy observed for high-frequency emotion words and reduced accuracy of lower-frequency, more nuanced labels. Language proficiency emerged as the strongest predictor of productive emotion vocabulary, while trait Emotionality showed a smaller but independent association with productive emotion vocabulary. Importantly, no interaction between proficiency and Emotionality was observed, suggesting that linguistic and emotional capacities contribute additively rather than interactively to emotion-word production.

Naming patterns revealed that all the items had responses matching the target emotions or close synonyms based on dominance scoring, confirming the robustness of the vignette design in evoking contextually appropriate emotional responses. These items have similar response patterns based on naming agreement (e.g., "inquisitive": "curious", "excite", and "inquisitive"; "dizzy": "dizzy", "tire", and "sick"). This indicates that the vignettes were well designed and largely able to elicit the target emotions from teachers. Beyond supporting the robustness of the task design, the strong naming agreement also reflects teachers' shared conceptual understanding of common emotional experiences.

Consistent with findings reported in Chee et al. (2025), several low frequency items have synonyms from higher frequency bands (e.g., inquisitive – curious, disbelief – doubt). Therefore, it is unsurprising that teachers actually provided synonyms from the higher frequency bands instead of the low frequency target emotion word. From a psycholinguistic perspective, this pattern may also reflect Zipf's Principle of Least Effort (Zipf, 2016), which posits that speakers tend to select linguistic forms that require the least cognitive effort to retrieve and produce. In this case, participants may have defaulted to more accessible, higher-frequency emotion terms even when more specific lexical alternatives were available. At the same time, it remains unclear whether this pattern is primarily driven by lexical frequency effects, limitations in participants' language proficiency, or a more complex interaction between the two that varies across items and individuals. While higher-frequency synonyms may represent efficient communicative choices, they may also signal constraints in the breadth or accessibility of teachers' affective lexicon. Conceptually, this pattern illustrates how teachers may rely on broader, more accessible terms ("sad", "angry") rather than nuanced emotions ("disbelief", "doubt"), potentially limiting their precision in describing and understanding learners' emotional states. For teachers and researchers, this finding suggests that the ability to retrieve and produce emotion vocabulary is constrained by lexical knowledge, further emphasising the need for explicit instruction and practice with less commonly used emotional terms in ESL/EFL curriculum.

Correlation analysis revealed that language proficiency had the highest correlation with the accuracy and modulated frequency scores, which reinforces the role of word frequency in emotion vocabulary

(Chen & Truscott, 2010; Liu et al., 2011; Randolph et al., 1999; Schmitt & Schmitt, 2014, 2020; Wilkens et al., 2014). This suggests that teachers with higher language proficiency are more likely to produce more specific and complex emotion vocabulary. Global trait EI showed a weak correlation accuracy scoring. Only emotionality showed a significant, albeit weak, correlation with accuracy and modulated frequency scores in the PEVST. This finding is consistent with the view that emotion perception and expression may be associated with understanding different emotional contexts and the use of emotion vocabulary (Barrett, 2017; Hoemann et al., 2019).

In addressing RQ2, the regression model explains a moderate proportion of variance (24%), indicating that both language proficiency and EI independently predict productive emotion vocabulary. LexTALE emerged as the strongest positive predictor, showing that lexical proficiency remains a core factor for expressing emotion vocabulary. The Emotionality factor (encompassing emotion perception and expression) showed a small, independent positive association with productive labelling, above proficiency and demographics. This supports the view that dispositions toward perceiving/expressing emotions modestly co-vary with access to affective labels, even when general lexical skill is accounted for (Barrett, 2017; Hoemann et al., 2019). Among demographic factors, in-service teachers outperformed pre-service teachers, However, this difference should be interpreted cautiously. The present study did not measure teaching experience, classroom exposure, or interpersonal interaction variables that might plausibly contribute to differences in affect-related language use. It is therefore unclear whether the observed effect reflects teaching experience, cohort and age-related differences, or other unmeasured factors. Future studies could examine these possibilities by including measures of teaching experience, classroom interaction frequency, or affect-related pedagogical training.

The negative association between age and accuracy was small and may be attributable to cohort differences or other unmeasured covariates; consequently, this finding should be interpreted with caution. No gender effects were observed in this sample; however, the gender distribution was unbalanced, hence this null finding should be interpreted cautiously (see Bazhydai et al., 2019; Dylman et al., 2020; Mavrou, 2021) suggesting that emotional and linguistic predictors are more dominant in accounting for variance in productive emotion vocabulary knowledge, rather than demographic characteristics. Importantly, the non-significant LexTALE × Emotionality interaction indicates additive rather than moderating effects. In short, within the present range of proficiency and EI, the skills appear complementary and largely independent contributors to productive affective labelling. This finding is consistent with previous studies suggesting that while EI supports the contextual understanding and use of emotion words (Barrett, 2017; Hoemann et al., 2019), language proficiency ultimately determines the accessibility and precision of lexical retrieval (Chen & Truscott, 2010; Schmitt & Schmitt, 2020).

Findings from this study align with calls for the curriculum to prioritise the development of teachers' and students EI and emotion vocabulary (Council of Europe, 2020; Ministry of Education Malaysia, 2013; New South Wales Government, 2018; Turner & Stough, 2020). The relationship between emotionality and emotion vocabulary further highlights the importance of integrating EI into language instruction, enabling teachers to design activities that bolster both linguistic proficiency and emotional awareness. The prevalence of high-frequency synonyms (e.g., "inquisitive – curious") highlights systemic gaps in ESL/EFL materials, which often neglect mid- and low-frequency emotion words, essential for developing advanced emotional competence (Sánchez & Pérez-García, 2020). From a pedagogical standpoint, this underscores the value of integrating emotional scenarios or narratives in classroom instruction to enhance both language and EI development. Interventions could focus on practising lower-frequency near-synonyms in contextualised feedback scripts (e.g., distinguishing regret/remorse/guilt; nostalgic/sentimental/longing). Since emotion vocabulary reflects not only lexical knowledge but also emotional awareness (Barrett, 2017), targeted instruction can cultivate both linguistic and socio-emotional competencies. Embedding such practices aligns with the broader educational agendas that advocate EI training and emotional literacy as part of holistic language education (Council of Europe, 2020; Turner & Stough, 2020). At the policy level, findings derived from this study can

potentially inform curriculum design and EI-focused interventions, contributing to emotionally literate and communicatively competent learners. In-service teachers' higher scores may reflect richer exposure to classroom affective discourse. Embedding practicum tasks that require precise affect labelling, feedback scripting, and peer review could narrow the gap for pre-service teachers.

Some limitations of the present study warrant cautious interpretations of the findings. The use of snowball sampling resulted in a sample that was skewed toward upper-intermediate and advanced proficiency levels, as well as a notable over-representation of female participants. Consequently, the results may not fully generalise to the broader teacher population, particularly those with lower English proficiency. Future studies should adopt a more targeted recruitment strategy to ensure a more balanced representation across proficiency levels, gender, and teaching contexts, which would enhance response variability and representativeness. Additionally, the cross-sectional design of this study supports only associational, rather than causal, interpretations of the relationships. The small negative association between age and emotion vocabulary accuracy should also be interpreted cautiously, as it may reflect cohort or training-related differences.

6 Conclusion

This study examined the relationship between teachers' EI and productive emotion vocabulary within ESL/EFL contexts, using the PEVST as a novel measure of productive emotion vocabulary knowledge. Using the PEVST, a vignette-based measure of productive affect-related labelling, we found that teachers performed strongly on high-frequency items but showed reduced accuracy for several mid- and lower-frequency targets, indicating robust frequency constraints on productive access (RQ1). In regression analyses, LexTALE emerged as the strongest predictor of PEVST accuracy, with trait Emotionality showing a smaller independent positive association, and no evidence of LexTALE-by-Emotionality moderation (RQ2). These findings motivate greater attention in teacher education and materials to contextualised practice with lower-frequency affective labels, while future work should test whether targeted instruction improves productive access and whether these patterns generalise across more diverse teacher populations. By highlighting the link between emotionality and emotion vocabulary, this study contributes to the growing recognition that effective teaching is not only a linguistic role but also an emotional one.

Appendix A: PEVST items

PEVST items

Item	Target emotion	Frequency	Valence	Arousal
Q1	happy	1k	8.48	6.50
Q2	regret	3k	3.41	4.90
Q3	embarrass	2k	3.38	5.47
Q4	disgust	2k	3.32	5.00
Q5	lone	2k	3.75	3.84
Q6	worry	1k	3.27	5.81
Q7	safe	1k	7.70	3.14
Q8	jealous	5k	2.38	5.90
Q9	surprise	1k	7.44	6.57
Q10	boring	1k	2.95	3.65
Q11	excite	1k	5.68	5.43
Q12	lazy	1k	3.90	2.76
Q13	obsess	4k	3.23	4.95

Q14	dizzy	5k	3.36	4.95
Q15	relieve	3k	7.25	3.90
Q16	nostalgia	6k	6.65	4.38
Q17	shy	1k	5.16	3.33
Q18	greed	4k	2.48	4.45
Q19	disbelief	12k	4.21	5.58
Q20	hungry	1k	3.20	4.80
Q21	curious	2k	6.37	5.90
Q22	impulse	4k	5.16	5.33
Q23	grateful	3k	7.50	4.29
Q24	flirt	6k	6.73	5.93
Q25	inquisitive	9k	6.00	5.33

References

Alam, A., & Ahmad, M. (2018). The role of teachers' emotional intelligence in enhancing student achievement. *Journal of Asia Business Studies*, *12*(1), 31–43. https://doi.org/10.1108/JABS-08-2015-0134

Aldrup, K., Carstensen, B., & Klusmann, U. (2024). *The role of teachers' emotion regulation in teaching effectiveness: A systematic review integrating four lines of research*. https://doi.org/10.1080/00461520.2023.2282446

Anari, N. N. (2012). Teachers: emotional intelligence, job satisfaction, and organizational commitment. *Journal of Workplace Learning*, *24*(4), 256–269. https://doi.org/10.1108/13665621211223379

Arens, A., & Morin, A. J. S. (2016). Relations between teachers' emotional exhaustion and students' educational outcomes. *Journal of Educational Psychology*, *108*(6), 800–813. https://doi.org/10.1037/edu0000105

Aryadoust, V., Ng, L. Y., & Sayama, H. (2021). A comprehensive review of Rasch measurement in language assessment: Recommendations and guidelines for research. *Language Testing*, *38*(1), 6–40. https://doi.org/10.1177/0265532220927487/ASSET/IMAGES/LARGE/10.1177_0265532220927487-FIG2.JPEG

Barrett, L. F. (2017). *How emotions are made: The secret life of the brain*. Pan Macmillan.

Bazhydai, M., Ivcevic, Z., Brackett, M. A., & Widen, S. C. (2019). Breadth of Emotion Vocabulary in Early Adolescence. *Imagination, Cognition and Personality*, *38*(4), 378–404. https://doi.org/10.1177/0276236618765403

Bielak, J., & Mystkowska-Wiertelak, A. (2022). Language teachers' interpersonal learner-directed emotion-regulation strategies. *Language Teaching Research*, *26*(6), 1082–1105. https://doi.org/10.1177/1362168820912352/FORMAT/EPUB

Bond, T. (2015). Applying the Rasch Model : Fundamental Measurement in the Human Sciences, Third Edition. *Applying the Rasch Model*. https://doi.org/10.4324/9781315814698

Bonett, D. G., & Wright, T. A. (2015). Cronbach's alpha reliability: Interval estimation, hypothesis testing, and sample size planning. *Journal of Organizational Behavior*, *36*(1), 3–15. https://doi.org/10.1002/JOB.1960

Boone, W. J. (2016). Rasch Analysis for Instrument Development: Why, When, and How? *CBE Life Sciences Education*, *15*(4). https://doi.org/10.1187/CBE.16-04-0148

Boone, W. J., Yale, M. S., & Staver, J. R. (2014). Rasch analysis in the human sciences. *Rasch Analysis in the Human Sciences*, 1–482. https://doi.org/10.1007/978-94-007-6857-4/COVER

Brackett, M., & Cipriano, C. (2020). Emotional Intelligence Comes of Age. *Cerebrum: The Dana Forum on Brain Science*. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7409775/

Brown, H. D. (2014). Language learning and teaching. In *Language Teaching* (6th editio). Pearson Education. https://doi.org/10.1017/S0261444800011472

Chang, M. L. (2020). Emotion Display Rules, Emotion Regulation, and Teacher Burnout. *Frontiers in Education*, *5*, 536733. https://doi.org/10.3389/FEDUC.2020.00090/BIBTEX

Chee, A. J. E., Szabo, C. Z., & Ambrose, S. (2025). Measuring Emotion Recognition Through Language: The Development and Validation of an English Productive Emotion Vocabulary Size Test. *Languages 2025, Vol. 10, Page 204*, *10*(9), 204. https://doi.org/10.3390/LANGUAGES10090204

Chen, C., & Truscott, J. (2010). The Effects of Repetition and L1 Lexicalization on Incidental Vocabulary Acquisition. *Applied Linguistics*, *31*(5), 693–713. https://doi.org/10.1093/APPLIN/AMQ031

Chen, J., & Guo, W. (2020). Emotional intelligence can make a difference: The impact of principals' emotional intelligence on teaching strategy mediated by instructional leadership. *Educational Management Administration & Leadership*, *48*(1), 82–105.

Cheng, X., Xie, H., Hong, J., Bao, G., & Liu, Z. (2022). Teacher's Emotional Display Affects Students' Perceptions of Teacher's Competence, Feelings, and Productivity in Online Small-Group Discussions. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/FPSYG.2021.795708/FULL

Corcoran, R. P., & O'Flaherty, J. (2022). Social and emotional learning in teacher preparation: Pre-service teacher well-being. *Teaching and Teacher Education*, *110*, 103563.

Council of Europe. (2020). *Common European framework of reference for languages: Learning, teaching, assessment companion volume*. Council of Europe Publishing. www.coe.int/lang-cefr

Curci, A., Lanciano, T., & Soleti, E. (2014). Emotions in the classroom: The role of teachers' emotional intelligence ability in predicting students' achievement. *American Journal of Psychology*, *127*(4), 431–445. https://doi.org/10.5406/amerjpsyc.127.4.0431

Dewaele, J.-M., Gkonou, C., & Mercer, S. (2018). Do ESL/EFL Teachers' Emotional Intelligence, Teaching Experience, Proficiency and Gender Affect Their Classroom Practice? In J. de D. M. Agudo (Ed.), *Emotions in Second Language Teaching* (pp. 125–141). Springer International Publishing. https://doi.org/10.1007/978-3-319-75438-3_8

Dylman, A. S., Blomqvist, E., & Champoux-Larsson, M. F. (2020). Reading habits and emotional vocabulary in adolescents. *Educational Psychology*, *40*(6), 681–694. https://doi.org/10.1080/01443410.2020.1732874

Ebert, M., Ivcevic, Z., Widen, S. S., Linke, L., & Brackett, M. (2014). *Breadth of Emotion Vocabulary in Middle Schoolers*. http://elischolar.library.yale.edu/dayofdatahttp://elischolar.library.yale.edu/dayofdata/2014/Posters/6

Encyclopædia Britannica. (2024). *Find Definitions & Meanings of Words | Britannica Dictionary*. https://www.britannica.com/dictionary

Fisher, W. P. J. (1992). Reliability statistics. *Rasch Measurement Transactions*, *6*, 238.

Frenzel, A. C., Daniels, L., & Burić, I. (2021). Teacher emotions in the classroom and their implications for students. *Educational Psychologist*, *56*(4), 250–264. https://doi.org/10.1080/00461520.2021.1985501

Gaies, S. J. (1983). The investigation of language classroom processes. *TESOL Quarterly*, *17*(2), 205–217.

Gallingane, C., & Han, H. S. (2015). Words Can Help Manage Emotions: Using Research-Based Strategies for Vocabulary Instruction to Teach Emotion Words to Young Children. *Childhood Education*, *91*(5), 351–362. https://doi.org/10.1080/00094056.2015.1090849

Gass, S. M., & Madden, C. G. (1985). *Input in Second Language Acquisition.* ERIC.

Gkonou, C., Dewaele, J.-M., & King, J. (2020). Introduction to the Emotional Rollercoaster of Language Teaching [Book]. In C. Gkonou, J.-M. Dewaele, & J. King (Eds.), *The Emotional Rollercoaster of Language Teaching* (pp. 1–12). Multilingual Matters.

Gkonou, C., & Miller, E. R. (2019). Caring and emotional labour: Language teachers' engagement with anxious learners in private language school classrooms. *Language Teaching Research*, *23*(3), 372–387. https://doi.org/10.1177/1362168817728739

Gkonou, C., & Miller, E. R. (2020). *An Exploration of Language Teacher Reflection, Emotion Labor, and Emotional Capital*. https://doi.org/10.1002/tesq.580

Glas, C. A. W., & Verhelst, N. D. (1995). Testing the Rasch Model. In G. H. Fischer & I. W. Molenaar (Eds.), *Rasch Models: Foundations, Recent Developments, and Applications* (pp. 69–95). Springer New York. https://doi.org/10.1007/978-1-4612-4230-7_5

Hargreaves, A. (2001). Emotional geographies of teaching. *Teachers College Record*, *103*(6), 1056–1080. https://doi.org/10.1111/0161-4681.00142

Hoemann, K., Xu, F., & Barrett, L. F. (2019). Emotion words, emotion concepts, and emotional development in children: A constructionist hypothesis. *Developmental Psychology*, *55*(9), 1830–1849. https://doi.org/10.1037/dev0000686

Humphrey, N., Barlow, A., Wigelsworth, M., Lendrum, A., Pert, K., Joyce, C., Stephens, E., Wo, L., Squires, G., Woods, K., Calam, R., Harrison, M., & Turner, A. (2015). Promoting Alternative Thinking Strategies (PATHS) Evaluation report and Executive summary. In *Education Endownment Foundation* (Number June). www.educationendowmentfoundation.org.uk

Isenbarger, L., & Zembylas, M. (2006). The emotional labour of caring in teaching. *Teaching and Teacher Education*, *22*(1), 120–134. https://doi.org/10.1016/j.tate.2005.07.002

Jennings, P. A., & Greenberg, M. T. (2009). The Prosocial Classroom: Teacher Social and Emotional Competence in Relation to Student and Classroom Outcomes. *Review of Educational Research*, *79*(1), 491–525. https://doi.org/10.3102/0034654308325693

King, J. (2016). "It's Time, Put on the Smile, It's Time!": The Emotional Labour of Second Language Teaching Within a Japanese University. In Christina Gkonou, D. Tatzl, & S. Mercer (Eds.), *New Directions in Language Learning Psychology* (pp. 97–112). Springer, Cham. https://doi.org/10.1007/978-3-319-23491-5_7

Lee, Y. H., Kwon, H. H., & Richards, K. A. (2019). Emotional Intelligence, Unpleasant Emotions, Emotional Exhaustion, and Job Satisfaction in Physical Education Teaching. *Journal of Teaching in Physical Education*, 1–31. https://doi.org/10.1123/jtpe.2018-0177

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English. *Behavior Research Methods*, *44*(2), 325. https://doi.org/10.3758/S13428-011-0146-0

Linacre, J. M. (2024). A user's guide to WINSTEPS® MINISTEP: Rasch-model computer programs. *Program Manual 5.8.0*, 719.

Liu, D., & Zhong, S. (2016). L2 vs. L1 use of synonymy: An empirical study of synonym use/acquisition. *Applied Linguistics*, *37*(2), 239–261. https://doi.org/10.1093/APPLIN/AMU022

Liu, Y., Hao, M., Li, P., & Shu, H. (2011). Timed Picture Naming Norms for Mandarin Chinese. *PLOS ONE*, *6*(1), e16505. https://doi.org/10.1371/JOURNAL.PONE.0016505

Loh, C. E., & Liew, W. M. (2016). Voices from the ground: The emotional labour of English teachers' work. *Teaching and Teacher Education*, *55*, 267–278. https://doi.org/10.1016/j.tate.2016.01.016

Long, M. H. (2020). Optimal input for language learning: Genuine, simplified, elaborated, or modified elaborated? *Language Teaching*, *53*(2), 169–182.

Lüdecke, D., Ben-Shachar, M., Patil, I., Waggoner, P., & Makowski, D. (2021). performance: An R Package for Assessment, Comparison and Testing of Statistical Models. *Journal of Open Source Software*, *6*(60), 3139. https://doi.org/10.21105/joss.03139

Mavrou, I. (2021). Emotional intelligence, working memory, and emotional vocabulary in L1 and L2: Interactions and dissociations. *Lingua*, *257*. https://doi.org/10.1016/j.lingua.2021.103083

Mayer, J. D., Caruso, D. R., & Salovey, P. (2016). The Ability Model of Emotional Intelligence: Principles and Updates. *Emotion Review*, *8*(4), 290–300. https://doi.org/10.1177/1754073916639667

Mérida-López, S., & Extremera, N. (2017). Emotional intelligence and teacher burnout: A systematic review. *International Journal of Educational Research*, *85*, 121–130. https://doi.org/10.1016/j.ijer.2017.07.006

Merriam-Webster. (2024). *Merriam-Webster: America's Most Trusted Dictionary*. https://www.merriam-webster.com/

Ministry of Education Malaysia. (2013). *MALAYSIA EDUCATION BLUEPRINT 2013-2025*. Kementerian Pendidikan Malaysia. https://www.moe.gov.my/menumedia/media-cetak/penerbitan/dasar/1207-malaysia-education-blueprint-2013-2025/file

Ministry of Education Malaysia. (2019). *English Language Management Guidebook (Buku Panduan Bahasa Inggeris)*. The Inspectorate. https://www.moe.gov.my/en/muat-turun/penerbitan-dan-jurnal/buku-panduan/3888-buku-panduan-bahasa-inggeris/file

Molero, P. P., Ortega, F. Z., Jiménez, J. L. U., & Valero, G. G. (2019). Influence of Emotional Intelligence and Burnout Syndrome on Teachers Well-Being: A Systematic Review. *Social Sciences 2019, Vol. 8, Page 185*, *8*(6), 185. https://doi.org/10.3390/SOCSCI8060185

Nation, P. (2004). A study of the most frequent word families in the British National Corpus. *Vocabulary in a Second Language*, 3–13.

New South Wales Government. (2018). *New standards for NSW's teachers*. https://www.nsw.gov.au/news/new-standards-for-nsws-teachers

Nyanjom, J., & Naylor, D. (2021). Performing emotional labour while teaching online. *Educational Research*, *63*(2), 147–163. https://doi.org/10.1080/00131881.2020.1836989

Pavlenko, A. (2008). Emotion and emotion-laden words in the bilingual lexicon. *Bilingualism*, *11*(2), 147–164. https://doi.org/10.1017/S1366728908003283

Petrides, K. V. (2009). *Psychometric Properties of the Trait Emotional Intelligence Questionnaire (TEIQue)*. 85–101. https://doi.org/10.1007/978-0-387-88370-0_5

Petrides, K. V, & Mavroveli, S. (2018). Theory and applications of trait emotional intelligence. *Psychology: The Journal of the Hellenic Psychological Society*, *23*(1), 24–36.

Petrides, K. V, Sanchez-Ruiz, M.-J., Siegling, A. B., Saklofske, D. H., & Mavroveli, S. (2018). *Emotional Intelligence as Personality: Measurement and Role of Trait Emotional Intelligence in Educational Contexts* (pp. 49–81). Springer International Publishing AG. https://doi.org/10.1007/978-3-319-90633-1_3

Pourbahram, R., & Hajizadeh, M. (2018). The relationship between EFL instructors' emotional intelligence and learners' academic achievement. In *Language Teaching and Educational Research* (Number 1). http://dergipark.gov.tr/later

Pugh, E. V. (2008). Recognising Emotional Intelligence in Professional Standards for Teaching. *Practitioner Research in Higher Education*, *2*(1), 3–12.

Quintanilla, A., & Kloss Medina, S. (2024). Academic lexicon development in an EMI context: A study of pre-service English teachers' lexical availability. *International Journal of TESOL Studies, 6*(2), 61-84. https://doi.org/10.58304/ijts.20240205

Randolph, C., Lansing, A. E., Ivnik, R. J., Cullum, C. M., & Hermann, B. P. (1999). Determinants of Confrontation Naming Performance. *Archives of Clinical Neuropsychology*, *14*(6), 489–496. https://doi.org/10.1016/S0887-6177(98)00023-7

Revelle, W. (2023). *Procedures for Psychological, Psychometric, and Personality Research [R package psych version 2.3.12]*. https://CRAN.R-project.org/package=psych

Richards, J. C. (2022). Exploring Emotions in Language Teaching. *RELC Journal*. https://doi.org/10.1177/0033688220927531

RStudio Team. (2020). *RStudio: Integrated Development Environment for R* (4.3.3). RStudio, PBC.

Sánchez, M. J., & Pérez-García, E. (2020). Emotion(less) textbooks? An investigation into the affective lexical content of EFL textbooks. *System*, *93*, 102299. https://doi.org/10.1016/j.system.2020.102299

Schmitt, N., & Schmitt, D. (2014). A reassessment of frequency and vocabulary size in L2 vocabulary teaching1. *Language Teaching*, *47*(4), 484–503. https://doi.org/10.1017/S0261444812000018

Schmitt, N., & Schmitt, D. (2020). *Vocabulary in language teaching*. Cambridge University Press.

Schonert-Reichl, K. A. (2017). Social and Emotional Learning and Teachers. In *Social and Emotional Learning and Teachers* (Vol. 27, Number 1). www.futureofchildren.org

Schonert-Reichl, K. A., Jennifer Kitil, M., Hanson-Peterson, J., Jaramillo, A., Joosse, S., LeRose, M., Norman, N., Sipl, M., Sweiss, L., Teja, Z., Whitehead, J., Tyler Binfet, J., Donahue-Keegan, D., Jennings, P., Markowitz, N., Stillman, S., Wanless, S. B., Karen Niemi, B., & Weissberg, R. P. (2017). *To reach students, teach the teacher: A national scan of teacher preparation and social and emotional learning* (Number October). http://www.casel.org/wp-content/uploads/2017/02/SEL-TEd-Full-Report-for-CASEL-2017-02-14-R1.pdf

Smith, A. B., Rush, R., Fallowfield, L. J., Velikova, G., & Sharpe, M. (2008). *Rasch fit statistics and sample size considerations for polytomous data*. https://doi.org/10.1186/1471-2288-8-33

Smith Jr, E. V. (2002). Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. *Journal of Applied Measurement*, *3*(2), 205–231.

Song, J. (2018). Critical Approaches to Emotions of Non-Native English Speaking Teachers. *Chinese Journal of Applied Linguistics*, *41*(4), 453–467. https://doi.org/10.1515/cjal-2018-0033

Stanley, D. (2021). *Create American Psychological Association (APA) Style Tables [R package apaTables version 2.0.8]*. https://CRAN.R-project.org/package=apaTables

Szabo, C. Z., Bong, W. L. J., Wang, Y., Chee, A., & Lee, S. T. (forthcoming). Developing a productive emotion vocabulary test for adult speakers of english as an additional language. [Manuscript in preparation]

Taber, K. S. (2018). The Use of Cronbach's Alpha When Developing and Reporting Research Instruments in Science Education. *Research in Science Education*, *48*(6), 1273–1296. https://doi.org/10.1007/S11165-016-9602-2/TABLES/1

Takšić, V., & Mohorić, T. (2008). Vocabulary Emotion Test (VET): Ability measure of emotional intelligence. *XXIX International Congress in Psychology*, 423.

Taxer, J. L., & Frenzel, A. C. (2015). Facets of teachers' emotional lives: A quantitative investigation of teachers' genuine, faked, and hidden emotions. *Teaching and Teacher Education*, *49*, 78–88. https://doi.org/10.1016/J.TATE.2015.03.003

Teng, M. F. (2017). Emotional Development and Construction of Teacher Identity: Narrative Interactions about the Pre-service Teachers' Practicum Experiences. *Australian Journal of Teacher Education*, *42*(11), 42. http://ro.ecu.edu.au/ajte/vol42/iss11/8

Thomas, N. S., Mousavy, S., Mukundan, J., & Nimehchisalem, V. (2012). The Relationship between Burnout and Trait Emotional Intelligence among Secondary School Teachers in Malaysia. *International Journal of Applied Linguistics and English Literature*, *1*(4), 15–23. https://doi.org/10.7575/IJALEL.V.1N.4P.15

Turner, K., & Stough, C. (2020). Pre-service teachers and emotional intelligence: a scoping review. *Australian Educational Researcher*, *47*(2), 283–305. https://doi.org/10.1007/s13384-019-00352-0

Valente, S., Veiga-Branco, A., Rebelo, H., Lourenço, A. A., & Cristóvão, A. M. (2020). The Relationship between Emotional Intelligence Ability and Teacher Efficacy. *Universal Journal of Educational Research*, *8*(3), 916–923. https://doi.org/10.13189/ujer.2020.080324

Valiente, C., Swanson, J., Delay, D., Fraser, A. M., & Parker, J. H. (2020). *Developmental Psychology Emotion-Related Socialization in the Classroom: Considering the Roles of Teachers, Peers, and the Classroom Context*. https://doi.org/10.1037/dev0000863

van Heuven, W. J. B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology*, *67*(6), 1176–1190. https://doi.org/10.1080/17470218.2013.850521/ASSET/IMAGES/LARGE/10.1080_17470218.2013.850521-FIG1.JPEG

Vince, R. (2016). Emotion and learning. *Journal of Management Education*, *40*(5), 538–544.

Wang, L. (2022). Exploring the relationship among teacher emotional intelligence, work engagement, teacher self-efficacy, and student academic achievement: A moderated mediation model. *Frontiers in Psychology*, *12*, 810559.

Wang, Y. (2022). To Be Expressive or Not: The Role of Teachers' Emotions in Students' Learning. *Frontiers in Psychology*, *12*, 737310. https://doi.org/10.3389/FPSYG.2021.737310/BIBTEX

Wickham, H., Averick, M., Bryan, J., Chang, W., D', L., Mcgowan, A., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Lin Pedersen, T., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., … Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/JOSS.01686

Wilkens, R., Dalla Vecchia, A., Boito, M. Z., Padró, M., & Villavicencio, A. (2014). Size does not matter. Frequency does. A study of features for measuring lexical complexity. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *8864*, 129–140. https://doi.org/10.1007/978-3-319-12027-0_11/COVER

Wright, B. D. (1994). Reasonable mean-square fit values. *Rasch Meas Transac*, *8*, 370.

Yoo, J., & Carter, D. (2017). Teacher Emotion and Learning as Praxis: Professional Teacher Emotion and Learning as Praxis: Professional Development that Matters Development that Matters. *Australian Journal of Teacher Education*, *42*. https://doi.org/10.14221/ajte.2017v42n3.3

Zembylas, M. (2007). Emotional capital and education: Theoretical insights from Bourdieu. *British Journal of Educational Studies*, *55*(4), 443–463. https://doi.org/10.1111/j.1467-8527.2007.00390.x

Zhu, Y., Zhang, B., Wang, Q. A., Li, W., & Cai, X. (2018). The principle of least effort and Zipf distribution. *Journal of Physics: Conference Series*, *1113*(1), 012007. https://doi.org/10.1088/1742-6596/1113/1/012007

Zipf, G. K. (2016). *Human behavior and the principle of least effort: An introduction to human ecology*. Ravenio Books.

Allen Chee Jie Ein, Ph.D. is an Assistant Professor at the School of Education, University of Nottingham Malaysia. His research is situated within TESOL and focuses on emotional language and the role of emotional intelligence in second language use. He investigates how emotional expression varies across first and additional languages, particularly in multilingual contexts. Drawing on interdisciplinary perspectives, his research seeks to advance understanding of the affective dimensions of language learning, use, and education, with emerging interests in how these insights can inform socially and emotionally responsive language pedagogy.

Dr Csaba Z Szabo is a lecturer (assistant professor) in Language Education (TESOL) at the Institute for Language Education, Moray House School of Education and Sport, University of Edinburgh. His

research focuses on vocabulary learning, teaching, and assessment in bi- and multilingual contexts, with particular interests in emotion vocabulary, cross-linguistic influence, and academic achievement in EMI settings. He has led and collaborated on funded research projects, including work as principal investigator advancing the development and validation of the Productive Emotion Vocabulary Test (PEVT), evaluating cognate-based vocabulary resources, and validating the PTE-A. His work emphasises practical impact and evidence-informed teaching.

Sharimila Ambrose, Ph.D. is an Assistant Professor (Special & Inclusive Education) at the University of Nottingham Malaysia (UNM) and a qualified, certified Speech-Language Pathologist with teaching and clinical experience in various educational settings in the United States of America, New Zealand, Malaysia, and United Kingdom. Her research areas include examining language and learning development and difficulties that impact individuals from diverse backgrounds and finding ways to foster inclusion and cultural responsivity in classrooms/educational/intervention settings using pedagogy-based teaching and learning frameworks (ICF, UDL, etc.). Sharimila has presented several research-, academic- and community-based talks in Malaysia and internationally.

Supplementary Materials

Rasch Analysis

Both dichotomous and polytomous Rasch modelling were used for the dominance and accuracy scoring respectively. The dichotomous Rasch model assesses items with binary outcomes (i.e., correct or incorrect responses) where only the most dominant responses were used. The polytomous Rasch model handles items with multiple outcomes for the accuracy scoring where multiple responses were acceptable. Rasch modelling was used to evaluate the technical quality of the test by investigating the mean-square fit statistics (MSQ) (Boone, 2016). This analysis ensures that the items conform to the unidimensionality requirements of the Rasch model, which posits that all items should measure a single latent variable (Bond, 2015; Boone, 2016). For this study, weighted mean-square fit is used as it is more resistant towards sample size changes (see Aryadoust et al., 2021; Smith et al., 2008). Items with infit and outfit weighted mean-square fit statistics (MSQ) beyond the acceptable threshold of 0.7 – 1.3 (Boone, 2016; Smith Jr,2002; Wright, 1994) can be considered as not conforming to the unidimensionality requirements of the Rasch model and should be subjected to further scrutiny or consideration for removal.

Various reliability statistics such as Item separation reliability (ISR) and person separation reliability (PSR) were also calculated. PSR coefficient of 0.7 and above is considered as the minimum requirement for the test to be considered at an acceptable level to distinguish between low and high ability respondents (Fisher, 1992). However, a coefficient of 0.8 and above is required to be considered a good level of distinction (Linacre, 2024). Same scales apply for ISR. To investigate representativeness of the test, difficulty of each item was calculated and mapped onto the Wright map, which displays the distribution of item difficulties and person abilities. Wright map shows if the test has any flooring or ceiling effect. Flooring effect means that there are no items present for low-ability respondents, whereas ceiling effect means that there are no items present for high-ability respondents (Boone, 2016). This ensures that the items are neither too easy nor too difficult for the respondents (Boone et al., 2014). Wald's test was conducted to further assess item fit of the Rasch model (Glas & Verhelst, 1995).

Dichotomous Rasch model

Dominance scoring was used for the dichotomous Rasch model. To conduct dichotomous Rasch modelling, only the most dominant response for each item was considered correct, whereas non-dominant responses were considered incorrect. Out of 5,050 dominant responses, 2,537 responses were considered correct and marked as 1. Table S1 shows the fit statistics of the dichotomous Rasch model using only the most dominant response. Firstly, the technical quality of the PEVST using dominance scoring was inspecting for underfitting and overfitting items with infit and outfit MSQ between 0.7 – 1.3 (Boone, 2016; Smith et al., 2008; Wright, 1994). A t-test was also reported for infit and outfit, and the t-value ideally should fall within the range of -1.96 to 1.96 (Wright, 1994). Q13 displayed a slightly inflated outfit mean-square value, suggesting marginal overfit. However, given that its fit residuals were within tolerable limits and its discrimination index remained positive, the item was retained for subsequent analyses. Based on those criteria, there were no items flagged as misfitting. All items fall within the thresholds.

Table S1
Fit statistics of dichotomous Rasch model using only the most dominant response.

Item	Chisq	df	p-value	Outfit MSQ	Infit MSQ	Outfit t	Infit t	Discrim
Q1	105.701	100	0.329	1.047	1.063	0.510	0.985	0.200
Q2	108.244	100	0.269	1.072	1.070	0.507	0.652	0.192
Q3	88.593	100	0.786	0.877	0.919	-1.228	-1.119	0.427
Q4	82.738	100	0.895	0.819	0.903	-0.881	-0.638	0.414
Q5	113.553	100	0.167	1.124	1.104	0.911	1.033	0.114
Q6	80.233	100	0.927	0.794	0.851	-1.227	-1.215	0.506
Q7	84.058	100	0.874	0.832	0.890	-1.647	-1.471	0.450
Q8	81.678	100	0.909	0.809	0.860	-1.231	-1.226	0.492
Q9	112.085	100	0.192	1.110	1.093	0.864	0.990	0.150
Q10	97.994	100	0.538	0.970	1.023	-0.108	0.219	0.249
Q11	97.346	100	0.557	0.964	0.926	-0.144	-0.544	0.387
Q12	95.910	100	0.597	0.950	0.974	-0.369	-0.253	0.324
Q13	137.098	100	0.008	1.357	1.150	2.441	1.808	0.001
Q14	89.510	100	0.765	0.886	0.911	-1.195	-1.342	0.425
Q15	78.019	100	0.949	0.772	0.891	-1.205	-0.763	0.459
Q16	105.804	100	0.326	1.048	0.958	0.394	-0.501	0.255
Q17	97.639	100	0.548	0.967	0.984	-0.314	-0.225	0.308
Q18	107.635	100	0.283	1.066	1.086	0.497	1.005	0.157
Q19	97.778	100	0.544	0.968	0.983	-0.116	-0.116	0.216
Q20	103.519	100	0.385	1.025	1.035	0.230	0.388	0.227
Q21	86.076	100	0.838	0.852	0.934	-0.455	-0.268	0.314
Q22	103.782	100	0.378	1.028	0.992	0.191	0.011	0.122
Q23	100.210	100	0.475	0.992	0.950	-0.040	-0.757	0.321
Q24	103.110	100	0.396	1.021	1.024	0.171	0.243	0.185
Q25	101.338	100	0.444	1.003	0.951	0.116	-0.151	0.304

Item separation reliability (*ISR* = 0.88) shows that this model is adequate in terms of size and difficulty to assess respondents' emotion vocabulary production. However, person separation reliability (*PSR* = 0.61) indicates low discriminative power (Fisher, 1992; Linacre, 2024), therefore not able to properly discern between low and high ability respondents. This could be due to the restricted variance in participants' proficiency levels.

Figure S1 presents a Wright map of the dichotomous model. Majority of participants are clustered around 0 to 1 on the latent dimension and only a few participants were closer to the extreme ends of -2 and +2, suggesting that the sample has an above average ability. There were items present in all ranges of the logits; however, majority of the items are clustered between -1 to 0. There may be a potential flooring effect as participants with very low abilities may struggle with the test when scored with dominance scoring.

Figure S1

Wright map of person ability and item map for dichotomous Rasch model

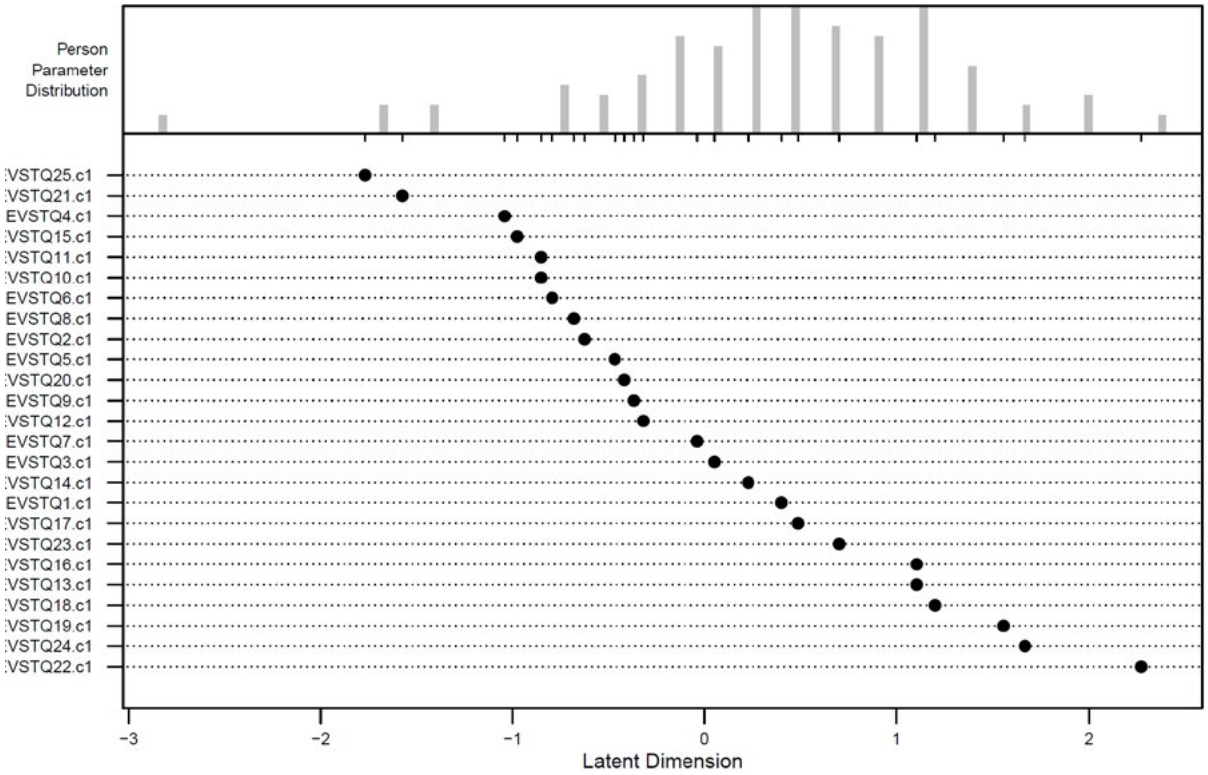


Table S2 shows a summary of Wald's test for the dichotomous Rasch model. There were no misfitting items as all items were within the threshold.

Table S2

Wald's test for dichotomous Rasch model.

Item	z-statistics	p-value
Q1	0.859	0.391
Q2	-0.627	0.531
Q3	0.076	0.939
Q4	-0.767	0.443
Q5	1.905	0.057
Q6	-0.741	0.459
Q7	-1.389	0.165
Q8	0.047	0.962
Q9	1.564	0.118
Q10	0.520	0.603
Q11	-0.603	0.547
Q12	-0.980	0.327
Q13	1.118	0.264
Q14	-0.585	0.559
Q15	-0.899	0.369
Q16	-0.191	0.848
Q17	-0.348	0.728

Q18	0.674	0.500
Q19	0.803	0.422
Q20	1.258	0.209
Q21	-0.447	0.655
Q22	0.639	0.523
Q23	-1.706	0.088
Q24	1.720	0.085
Q25	-0.135	0.892

Polytomous Rasch model

Accuracy scoring was adopted for the polytomous Rasch model. Table S3 shows the fit statistics of the polytomous Rasch model. All items appear to fall within the reasonable range.

Table S3
Fit statistics of polytomous Rasch model using accuracy scoring.

Item	Chisq	df	p-value	Outfit MSQ	Infit MSQ	Outfit t	Infit t	Discrim
Q1	115.419	100	0.139	1.143	0.965	0.612	-0.118	0.139
Q2	78.687	100	0.943	0.779	0.905	-0.789	-0.352	0.287
Q3	85.297	100	0.853	0.845	0.915	-0.928	-0.570	0.373
Q4	94.378	100	0.640	0.934	0.951	-0.401	-0.316	0.269
Q5	94.992	100	0.623	0.941	0.999	-0.146	0.068	0.150
Q6	88.097	100	0.797	0.872	0.977	-0.393	-0.018	0.145
Q7	75.147	100	0.970	0.744	0.922	-0.322	-0.028	0.150
Q8	84.770	100	0.862	0.839	0.920	-0.630	-0.334	0.294
Q9	86.514	100	0.830	0.857	0.979	-0.293	0.028	0.136
Q10	102.475	100	0.413	1.015	0.979	0.144	-0.094	0.205
Q11	81.474	100	0.912	0.807	0.853	-1.430	-1.136	0.464
Q12	96.972	100	0.567	0.960	0.974	-0.300	-0.185	0.258
Q13	107.392	100	0.289	1.063	1.059	0.593	0.575	0.185
Q14	99.006	100	0.509	0.980	0.965	-0.126	-0.261	0.330
Q15	108.262	100	0.269	1.072	1.060	0.628	0.540	0.146
Q16	102.167	100	0.421	1.012	1.021	0.139	0.223	0.188
Q17	104.105	100	0.369	1.031	1.019	0.289	0.199	0.216
Q18	91.375	100	0.719	0.905	0.904	-0.869	-0.899	0.400
Q19	84.865	100	0.860	0.840	0.838	-1.525	-1.568	0.519
Q20	100.728	100	0.461	0.997	1.001	0.018	0.051	0.220
Q21	96.300	100	0.586	0.953	0.961	-0.353	-0.301	0.276
Q22	99.057	100	0.508	0.981	0.994	-0.137	-0.017	0.287
Q23	83.909	100	0.876	0.831	0.926	-0.738	-0.333	0.355
Q24	108.361	100	0.267	1.073	1.071	0.632	0.615	0.126
Q25	95.382	100	0.612	0.944	0.959	-0.437	-0.348	0.279

Figure S2 presents a Wright map of the polytomous model. Majority of the participants clustered between 0 to +2 which indicates that this sample has an above average ability. There appears to be a good alignment between the respondents' abilities and item difficulties as there were no observable flooring or ceiling effects. All items appear to be evenly distributed across the logits. However, there were two items (Q25 and Q21) with abnormally wide thresholds.

Figure S2
Wright map of person ability and item map for polytomous Rasch model

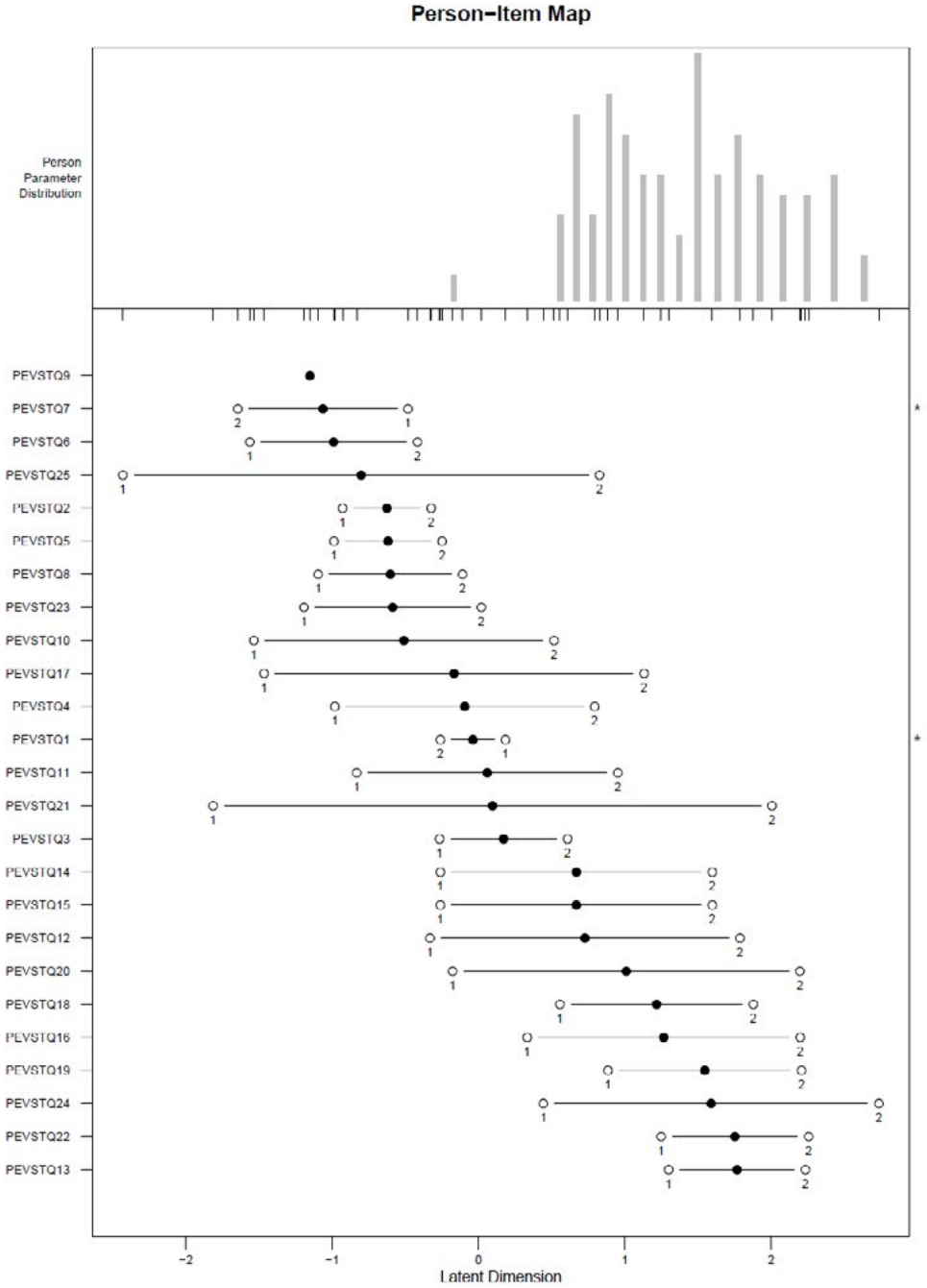


Table S4 shows a summary of the Wald's test for all 25 items. Wald tests could not be computed for eight items (Q2, Q4–Q8, Q10, Q11) due to sparse categories (empty/near-empty cells; no incorrect response); fit indices for these items were inspected and showed no substantive misfit.

Table S4
Wald's test for polytomous Rasch model

Items	z-statistic	p-value
Q3.c1	-0.116	0.908
Q3.c2	-0.634	0.526
Q9.c1	-0.425	0.671
Q12.c1	-0.527	0.598
Q12.c2	-0.711	0.477
Q13.c1	1.680	0.093
Q13.c2	0.419	0.675
Q14.c1	0.406	0.685
Q14.c2	-0.132	0.895
Q15.c1	-0.604	0.546
Q15.c2	-0.500	0.617
Q16.c1	-0.868	0.385
Q16.c2	0.035	0.972
Q17.c1	0.527	0.598
Q17.c2	0.459	0.646
Q18.c1	-0.418	0.676
Q18.c2	-1.249	0.212
Q19.c1	-1.510	0.131
Q19.c2	-1.482	0.138
Q20.c1	-0.559	0.576
Q20.c2	-0.522	0.602
Q21.c1	1.660	0.097
Q21.c2	1.313	0.189
Q22.c1	0.531	0.596
Q22.c2	-0.007	0.994
Q24.c1	0.132	0.895
Q24.c2	1.453	0.146