JIGE 6 (2) (2025) 1079-1100
JURNAL ILMIAH GLOBAL EDUCATION
ejournal.nusantaraglobal.ac.id/index.php/jige
DOI: https://doi.org/10.55681/jige.v6i2.3968

Non-Test Instrument for Pronunciation Self-Assessment among Intermediate-Level English Language Learners in Higher Education: A Rasch Model Approach

Syahadah Albaqiyatul Karimah[1*], Ninuk Lustyantie[2], Fathiaty Murtadho[2]
[1] Universitas Nahdlatul Ulama Indonesia
[2] Universitas Negeri Jakarta, Indonesia
*Corresponding author email: syahalbakarimah@unusia.ac.id

Article Info

Article history:
Received June 10, 2025
Approved June 28, 2025

Keywords:
English Language Learners, Intermediate Level, Non-Test Instrument, Pronunciation Self-Assessment, Rasch Model,

ABSTRACT

Pronunciation is a vital component of oral language proficiency, yet it remains underrepresented in English language instruction, particularly within higher education EFL/ESL contexts. This study aims to address this pedagogical gap by developing and validating a non-test self-assessment instrument to measure pronunciation self-efficacy among intermediate English language learners (ELLs). Grounded in constructivist and learner-centered theories, the instrument integrates four dimensions: Phonological Awareness (PA), Accent and Intonation Awareness (AIA), Self-Confidence (SC), and Self-Monitoring (SM). A total of 43 university students participated, selected via purposive sampling. The instrument, comprising 20 Likert-scale items, underwent content validation by three experts, followed by psychometric evaluation through Rasch analysis using Winsteps software. The results demonstrate strong psychometric properties across all dimensions. Item and person reliability values ranged from moderate to high (0.64–0.90), with Cronbach's Alpha (KR-20) between 0.68 and 0.90, indicating internal consistency. Person separation indices varied from 1.33 to 2.70, suggesting sufficient ability discrimination. Item fit statistics were within acceptable thresholds, supporting construct validity. Principal Components Analysis (PCA) of residuals confirmed unidimensionality, particularly for the SC construct, where Rasch explained variance reached 70.6%. PT-measure correlations further supported item discrimination quality. The validated instrument fills a methodological gap in the assessment of pronunciation self-efficacy and offers a practical, learner-centered tool for pedagogical implementation. It contributes to the advancement of self-regulated learning in pronunciation and opens pathways for further research into affective-cognitive correlates of pronunciation in EFL contexts.

Copyright © 2025, The Author(s).
This is an open access article under the CC–BY-SA license

How to cite: Karimah, S. A., Lustyantie, N., & Murtadho, F. (2025). Non-Test Instrument for Pronunciation Self-Assessment among Intermediate-Level English Language Learners in Higher Education: A Rasch Model Approach. *Jurnal Ilmiah Global Education*, *6*(2), 1079–1100. https://doi.org/10.55681/jige.v6i2.3968

INTRODUCTION

The introduction contains the purpose of article/research that is formulated and presented by an adequate introduction and avoids detail references and research result presentations. The research urgency, supporting facts, and data must be included. A preliminary research result should be explained as the basis of the research. Before mentioning the objective/s, a gap analysis must be elucidated. The gap analysis states the difference/s between the research and other previous studies. At this point, the novelty will be apparent. The research stance must be included, whether it corrects, debates, or support the previous research.

Pronunciation plays a pivotal role in effective oral communication and overall language intelligibility (Algethami & Al Kamli, 2025). For intermediate-level English language learners in higher education, accurate pronunciation is not only essential for academic success but also for future professional engagement in globalized contexts (Nguyen, 2024). Despite its importance, pronunciation often receives limited attention in the language assessment process, particularly in the form of self-assessment instruments that are reliable, valid, and tailored to learner proficiency levels.

In recent years, there has been a growing emphasis on learner autonomy and self-regulated learning in English language teaching. Self-assessment is recognized as a valuable tool to foster metacognitive awareness and learner reflection, particularly in pronunciation learning, where internal monitoring and repeated self-evaluation are key to improvement. However, many existing tools for assessing pronunciation are either test-based or not psychometrically evaluated, leading to issues of subjectivity, inaccuracy, or misalignment with the learner's actual competence.

The development of a non-test instrument for pronunciation self-assessment addresses this gap by offering a learner-centered, formative assessment alternative that emphasizes introspective judgment rather than performance in controlled settings. Furthermore, ensuring that the instrument adheres to rigorous measurement standards is critical for its credibility and usefulness. The Rasch Model, a modern psychometric approach rooted in item response theory (IRT), provides a robust framework for developing and validating such instruments. It offers the ability to calibrate item difficulty, assess unidimensionality, and ensure fairness across different learner groups.

Pronunciation is widely regarded as a pivotal component of oral language proficiency, particularly within the domains of English as a Foreign Language (EFL) and English as a Second Language (ESL) instruction. The ability to produce phonologically accurate and intelligible speech is not only integral to effective verbal communication but also to the development of broader linguistic competence (Derwing & Munro, 2005; Celce-Murcia et al., 2010; Nu My Nhat & Thi Thu Hien, 2024); Despite its essential role, pronunciation has traditionally received less instructional focus compared to grammar and vocabulary, often being relegated to incidental teaching (Levis, 2024). This pedagogical imbalance has hindered learners' communicative effectiveness, especially in high-stakes academic and professional contexts.

The constructs of intelligibility, comprehensibility, and communicative competence have been extensively explored in second language acquisition research (Derwing & Munro, 2022);(Kang & Hirschi, n.d.)). Intelligibility, defined as the extent to which a speaker's message is recognized by the listener, is distinct from comprehensibility, which pertains to the listener's ease of understanding. These two aspects, together with perceived accentedness, influence communicative success more than native-like pronunciation (Isaacs & Trofimovich, 2012).

Consequently, a pronunciation focus aimed at improving listener-oriented features rather than native norms aligns better with current pedagogical priorities.

Intermediate-level learners often exhibit persistent difficulties with both segmental and suprasegmental features of English. Segmental errors, such as mispronunciation of individual phonemes (e.g., /θ/ or /v/), can compromise word recognition, while suprasegmental inaccuracies involving rhythm, stress, and intonation disrupt discourse-level coherence (Hirschi & Kang, 2024)). The fossilization of such errors, particularly when feedback mechanisms are limited, further complicates pronunciation acquisition at this stage (Selinker, 2015); Albelihi & Al-Ahdal, 2024). These issues are compounded in EFL contexts where authentic exposure to English is limited, thereby constraining opportunities for naturalistic phonological input and output.

In addition to phonological challenges, affective variables such as pronunciation anxiety and low self-perceived intelligibility significantly inhibit oral communication. Learners frequently report feeling apprehensive about negative evaluation due to their accent or mispronunciations, which in turn diminishes their oral participation (Baran-Łucarz, 2014); Zega, 2025). These psychological constraints are particularly salient in pronunciation instruction, which inherently involves public speaking and auditory feedback. Research suggests that fostering pronunciation self-efficacy and providing supportive learning environments can mitigate these affective barriers and enhance learner engagement (Baranyi-Dupák, 2024).

Given the complex interplay of linguistic and affective factors, there is a pressing need for pedagogical models that empower learners to monitor and regulate their own pronunciation development. Recent scholarship advocates for learner-centered approaches that incorporate self-assessment, reflective learning, and formative feedback mechanisms (Regan et al., 2020). Such approaches not only promote autonomy but also align with current trends in second language assessment, which emphasize the role of learner perceptions and self-awareness in language learning trajectories. The development of non-test self-assessment instruments, when psychometrically validated using models such as Rasch analysis, represents a promising direction for pronunciation pedagogy in higher education settings.

Over the past two decades, self-assessment has emerged as a central strategy in fostering learner autonomy and enhancing metacognitive engagement in second language learning. Rooted in learner-centered pedagogies, self-assessment allows students to take an active role in evaluating their language abilities, thereby encouraging greater investment in the learning process (Raikhel, 2025). This reflective practice has been associated with increased self-regulation, deeper learning strategies, and improved academic outcomes (Zhai et al., 2023). In alignment with constructivist theory, self-assessment positions learners not merely as recipients of knowledge, but as active participants capable of shaping their linguistic development.

Research has consistently highlighted the multifaceted benefits of self-assessment, encompassing both cognitive and affective domains. Cognitively, it promotes metalinguistic awareness and the ability to identify specific language features that require attention (Hofer & Spechtenhauser, 2024). Affectively, self-assessment fosters motivation, confidence, and a sense of responsibility, which are critical for sustaining long-term engagement in language learning (Namaziandost et al., 2024). These benefits are particularly pronounced in adult and higher education learners, who often seek autonomy and relevance in their learning experiences (Yang et al., 2025). As such, integrating structured self-assessment practices into curricula is increasingly seen as a pedagogical imperative.

While self-assessment has been widely implemented in the assessment of reading, writing, listening, and speaking skills, its application in pronunciation instruction remains markedly limited (Zulkifly, 2023). The majority of existing instruments focus on general language proficiency, often neglecting the nuanced features of phonological performance. Moreover, pronunciation is frequently perceived as too technical or too dependent on expert evaluation to be self-monitored effectively (Zhu et al., 2024). This perception has contributed to the marginalization of pronunciation-focused self-assessment tools in formal language programs, despite evidence suggesting their potential utility.

The underutilization of pronunciation self-assessment may be attributed to several pedagogical and methodological barriers. From a pedagogical standpoint, instructors may lack training or resources to design valid and reliable self-assessment frameworks for pronunciation. Methodologically, the absence of standardized rubrics and empirical validation—especially within psychometric models like Rasch—limits the credibility of learner-generated assessments. Furthermore, learners themselves often struggle to evaluate their pronunciation objectively due to limited auditory discrimination or lack of linguistic terminology. These factors underscore the need for guided, scaffolded approaches that enhance learners' capacity to self-monitor phonological accuracy and fluency.

The evolving paradigm of language assessment has witnessed a growing interest in non-test instruments as viable complements—or alternatives—to conventional testing formats. Non-test tools, encompassing Likert-scale surveys, self-rating checklists, and reflective journals, foreground learners' introspective judgments about their language abilities (Lewkowicz, 2020). These instruments shift the emphasis from externally imposed measurement to internal self-regulation, aligning with contemporary pedagogical frameworks that value learner agency, formative feedback, and personalized learning trajectories. Within this reconceptualization, assessment is no longer solely a summative endpoint but an ongoing, learner-mediated process.

In the domain of pronunciation, non-test instruments are particularly valuable in capturing affective and perceptual dimensions of spoken language competence that formal tests may neglect. Traditional pronunciation assessments often prioritize phonetic precision and articulatory accuracy as judged by experts (Babaeian, 2023), yet such assessments may fail to reflect learners' actual communicative experiences. By contrast, self-report tools enable learners to evaluate comprehensibility, confidence, and functional fluency in real-life contexts (Plengkham, 2022). These dimensions are critical, especially for intermediate-level learners who may not yet have native-like accuracy but are capable of effective oral interaction.

Non-test instruments also serve as empowerment tools, enabling learners to assume evaluative responsibility for their own linguistic progress. Research has demonstrated that when learners are given structured opportunities to assess their pronunciation using self-rating scales, they develop heightened metacognitive awareness and self-monitoring capacity (Cai & Yu, 2024). This self-awareness, in turn, has been linked to improved oral proficiency outcomes, as learners can identify problem areas, seek targeted input, and engage in more strategic practice (Borg & Alshumaimeri, 2019). By decentralizing assessment from the instructor and integrating it into the learner's cognitive framework, non-test tools promote deeper engagement with pronunciation learning.

Despite their pedagogical promise, non-test instruments require rigorous methodological design to ensure validity, reliability, and interpretability. The use of psychometric models—

particularly the Rasch model—has been proposed as a robust means of calibrating item difficulty and rating scale functionality (T. Bond, 2015). In pronunciation self-assessment, Rasch analysis enables researchers to examine the fit between learners' self-perceptions and the theoretical construct of intelligibility or fluency, offering a nuanced profile of learner competence. Furthermore, item-person interaction mapping provides insight into the scalability and differentiation capacity of non-test tools, which is essential for their adoption in formal educational settings.

The integration of non-test instruments into pronunciation instruction holds considerable promise for enhancing feedback practices and promoting reflective learning. However, their effective deployment hinges on the development of clear rubrics, appropriate training, and culturally responsive design (Pan et al., 2024). As language learning becomes increasingly personalized and digitalized, non-test assessments are poised to play a pivotal role in learner-driven pronunciation development. Future research should continue to explore their alignment with communicative competence frameworks, their adaptability across proficiency levels, and their potential integration with digital and AI-mediated language learning environments.

In recent decades, there has been a growing emphasis on the need for psychometric instruments to undergo rigorous validation procedures to ensure their reliability, validity, and applicability across diverse populations. Traditional methods based on Classical Test Theory (CTT) have been widely used; however, their limitations—such as sample dependency and unequal measurement intervals—have prompted researchers to adopt more advanced approaches, such as Item Response Theory (IRT). Among IRT models, the Rasch Model has emerged as a particularly robust and widely adopted framework due to its strong theoretical foundation and practical utility in educational and psychological measurement (T. G. Bond & Fox, 2007;T. Bond, 2015). The model posits that the probability of a respondent endorsing an item is a logistic function of the difference between person ability and item difficulty, assuming a unidimensional latent trait.

The Rasch Model offers several advantages over traditional models, particularly in its capacity to convert ordinal raw scores into linear interval measures, thus allowing for more accurate statistical interpretations (Medvedev & Krägeloh, 2022). This transformation not only enhances measurement precision but also contributes to the overall construct validity of the instrument. Moreover, Rasch analysis facilitates the examination of item and person fit statistics, allowing researchers to identify anomalies such as misfitting items or unexpected response patterns. These diagnostics provide essential insights for refining and improving instrument quality. Dülger, 2020 underscores the model's capacity to ensure internal consistency, detect multidimensionality, and support the development of instruments that are fair and equitable across groups. As such, the Rasch Model is increasingly recognized as a gold standard in contemporary psychometric research, particularly when precision and diagnostic utility are paramount.

This study aims to develop and validate a non-test instrument for pronunciation self-assessment among intermediate-level English learners in higher education using the Rasch Model. By doing so, it not only contributes to the enhancement of pronunciation pedagogy but also provides empirical support for the integration of self-assessment tools into language learning curricula. Moreover, this research aligns with the broader educational trend toward personalized learning and learner empowerment, particularly within the context of 21st-century skills and Industry 5.0 demands.

METHODS

This study employed a quantitative research design to develop and validate a non-test instrument for pronunciation self-assessment among intermediate-level English language learners in higher education. The participants consisted of 43 university students from English language learning programs, all of whom were identified as having intermediate English proficiency based on institutional placement tests. A purposive sampling technique was used to ensure that the sample reflected the target population.

The instrument was constructed based on theoretical foundations from previous literature and comprised 20 Likert-scale items across four key dimensions: phonological awareness, self-monitoring, confidence, and accent-awareness. Expert judgment from three applied linguistics professionals was obtained to establish the content validity of the instrument. Minor revisions were made based on their feedback to improve item clarity and relevance. Data were collected through an online questionnaire distributed via Google Forms, ensuring ease of access for participants across multiple campuses. Informed consent was obtained, and confidentiality was maintained throughout the research process.

The Rasch model was applied using the Winsteps software to analyze the psychometric properties of the instrument. The analysis demonstrated satisfactory item and person reliability, appropriate item difficulty distribution, and acceptable fit statistics. In addition, Principal Component Analysis (PCA) of the residuals supported the unidimensionality of the scale. These findings confirmed that the developed instrument is valid and reliable for measuring pronunciation self-assessment among intermediate ELLs in higher education contexts.

Table 1. Design Phase

Steps	Activity	Method
1	Needs Analysis & Construct Definition	Literature Review
2	Expert Judgement (Pronunciation and Assessment)	Online discussion with 2 experts
3	Item Development	Item writing based on CEFR and pronunciation theory
4	Content Validation	Expert Review using Content Validity Index (CVI)
5	Pilot Testing & Data Collection	Administering instrument to target group using Likert-scale questionnaire
6	Rasch Analysis	Rasch modeling (fit analysis, unidimensionality), item refinement

The design phase of this study followed a systematic process to develop and validate a pronunciation assessment instrument. The first step involved a needs analysis and construct definition, which was conducted through an extensive literature review to identify relevant theories and frameworks, particularly those related to pronunciation assessment and the Common European Framework of Reference for Languages (CEFR). In the second step, expert judgement was sought through online discussions with two experts in pronunciation and assessment, aimed at refining the construct and ensuring theoretical alignment. The third step was item development, where test items were written based on CEFR descriptors and key principles from pronunciation theory. This was followed by a content validation phase, during which the items were reviewed by experts using the Content Validity Index (CVI) to assess relevance, clarity, and representativeness.

In the fifth step, pilot testing and data collection were conducted by administering the instrument to the target population using a Likert-scale questionnaire. Finally, Rasch analysis was employed to examine item fit, unidimensionality, and to perform item refinement, ensuring that the instrument met psychometric standards for validity and reliability.

RESULTS AND DISCUSSION

Descriptive Statistics and Rasch Summary

Table 2 presents the descriptive summary of Rasch measurement outcomes across four distinct dimensions assessed using five-item instruments administered to a sample of 43 respondents. For each dimension, descriptive Rasch analysis was applied to evaluate the psychometric adequacy of the scale, including mean raw scores, person measures in logits, standard errors, and variability in respondent ability. Key indicators such as person separation and person reliability (both real and model estimates) were reported to reflect the scale's capacity to differentiate among individuals. Fit statistics—namely Infit and Outfit Mean Square (MNSQ) values and standardized Z-scores (ZSTD)—were employed to assess the congruence between observed data and the Rasch model expectations. Internal consistency was evaluated using the Kuder-Richardson Formula 20 (KR-20), and the strength of association between raw scores and logit-based person measures was calculated to verify construct validity. Collectively, these results provide a comprehensive overview of measurement precision, reliability, and dimensional integrity across the examined constructs.

Table 2. Descriptive summary of Rasch Measurement

Dimension	N (Non-Extreme)	Raw Score Mean (SD)	Person Measure Mean (SE)	Person Separation	Person Reliability (Real/Model)	Infit MNSQ	Outfit MNSQ	ZSTD	KR-20 (α)	Raw–Logit Correlation
PA	43	19.2 (2.2)	1.62 (0.91)	1.70	0.67 / 0.74	0.98	0.99	-0.1	0.68	1.00
AIA	38	20.2 (3.0)	4.85 (1.80)	1.33	0.63 / 0.64	0.77	0.77	-0.3	0.90	0.96
Unnamed Set	42	16.4 (3.1)	1.08 (1.12)	2.70	0.88 / 0.90	0.88	0.89	~0	0.90	0.99
SM	40	—	1.72 (0.88)	1.57 / 1.91	0.71 / 0.79	0.99	0.99	~0	0.81	0.99

Descriptive Rasch analysis was conducted on 43 respondents across five items with four response categories. The average total raw score was 19.2 (SD = 2.2), indicating a relatively homogeneous distribution of responses. The mean person measure was 1.62 logits (SE = 0.91), reflecting adequate model precision in estimating individual abilities.

Fit statistics demonstrated satisfactory model-data alignment, with average Infit and Outfit MNSQ values of 0.98 and 0.99, respectively, and standardized Z-values near zero (ZSTD ≈ -0.1), supporting data normality and model conformity. Person reliability was 0.74 (model) and 0.67 (real), while Cronbach's Alpha (KR-20) reached 0.68, indicating moderate internal consistency. A person separation index of 1.70 suggested a reasonable ability to differentiate respondents based on their proficiency levels. The perfect correlation between raw scores and logit measures (r = 1.00) substantiated the construct validity of the instrument.

For the second dimension, data were obtained from the same 43 participants, with 38 classified as non-extreme. The average raw score was 20.2 (SD = 3.0), indicating slightly greater variability compared to the PA dimension. The average person measure was 4.85 logits with a higher standard error (SE = 1.80), pointing to increased measurement uncertainty.

Fit indices indicated acceptable conformity with the Rasch model, with Infit and Outfit MNSQ both at 0.77 and ZSTD values close to -0.3, suggesting a slight underfit but still within tolerable limits. Person reliability was 0.64 (model) and 0.63 (real), while KR-20 achieved a high level of internal consistency ($\alpha = 0.90$). The person separation index was 1.33, reflecting moderate differentiation capability. A strong correlation ($r = 0.96$) between raw scores and logit measures confirmed the dimensional integrity, although the relatively high error variance indicates room for enhancement in measurement accuracy.

The Rasch model was applied to data from 43 respondents and five items. Among the 42 non-extreme respondents, the average total raw score was 16.4 (SD = 3.1), corresponding to a mean person measure of 1.08 logits (SE = 1.12). The person separation index was 2.70, with a reliability of 0.88 for the real model and 0.90 for the expected model, indicating a good ability to discriminate between different ability levels. The correlation between raw score and Rasch person measure was remarkably high ($r = 0.99$), affirming strong internal validity of the measure.

Infit and outfit mean square statistics for the non-extreme persons were within acceptable thresholds (Infit MNSQ = 0.88; Outfit MNSQ = 0.89), suggesting a good model-data fit. The standard Z-scores were close to zero, indicating minimal distortion. The item reliability was moderate (0.70 for real, 0.74 for model), with a separation index of 1.53 and item mean measure centered at 0.00 logits. Cronbach's Alpha (KR-20) was 0.90, confirming high internal consistency of the instrument. However, improvement in item targeting may further optimize measurement precision.

For the SM dataset, Rasch analysis was conducted on responses from 43 participants and five items. Among 40 non-extreme cases, the average person measure was 1.72 logits (SE = 0.88), with a standard deviation of 1.90, indicating a moderate spread in respondent ability levels. The real person separation index was 1.57, and the person reliability was 0.71; the model-adjusted values were slightly higher (separation = 1.91; reliability = 0.79), suggesting satisfactory person differentiation but less than optimal for high-stakes measurement.

Fit statistics indicated an acceptable model fit (Infit MNSQ = 0.99; Outfit MNSQ = 0.99), and standardized residuals (ZSTD) were symmetrically distributed around zero, supporting the unidimensionality assumption. The correlation between raw scores and measures remained strong at 0.99. Cronbach's Alpha was calculated at 0.81, which is indicative of acceptable internal consistency.

The five items demonstrated good fit to the Rasch model (mean Infit MNSQ = 0.99), with a relatively narrow range of item difficulty (from -0.67 to 0.86 logits), suggesting that the test was well-targeted to the participants' ability levels. Nevertheless, further refinement of items could enhance separation and reliability.

Item Fit Statistics of the Overall Instrument

This section presents the results of the item fit analysis using Rasch modeling to evaluate the psychometric quality of the developed items across four dimensions: Phonological Awareness (PA), Accent and Intonation Awareness (AIA), Self-Confidence (SC), and Self-Monitoring (SM). Item fit statistics, including item difficulty, infit mean square (MNSQ), and point-measure correlation (PT-measure corr.), were examined to determine how well individual items aligned with the underlying constructs. These indicators are crucial in establishing the internal consistency, construct validity, and overall functionality of the instrument. Table 3 summarizes the key findings related to the range and distribution of item statistics across each dimension.

Table 3. Item Fit Statistics

Dimension	Item Difficulty Range (logits)	Infit MNSQ Range	Mean Infit	Item with Highest Infit	Item with Lowest Infit	PT-Measure Corr. Range	Conclusion
Phonological Awareness (PA)	-1.53 to 1.56	0.79 to 1.24	0.98	PA4 (1.24)	PA2 (0.79)	0.60 to 0.76	Very good and consistent
Accent and Intonation Awareness (AIA)	-0.77 to 0.39	0.85 to 1.49	0.99	AIA2 (1.49)	AIA3 (0.85)	0.75 to 0.85	Good overall, though AIA2 is slightly high
Self Confidence (SC)	Not reported	0.94 to 1.46	–	SC4 (1.46), SC1 (1.38)	SC3 (0.94)	0.65 to 0.78	Fairly good; minor revisions may be needed
Self Monitoring (SM)	-0.38 to 0.50	0.86 to 1.16	1.01	SM2 (1.16)	SM1 (0.86)	0.72 to 0.80	Consistent and construct-relevant

The results of the item fit analysis reveal that the dimensions of Phonological Awareness (PA) and Accent and Intonation Awareness (AIA) demonstrate sound psychometric properties. PA items exhibit good model-data fit, with infit MNSQ values ranging from 0.79 to 1.24 (M = 0.98), and PT-measure correlations between 0.60 and 0.76, suggesting moderate to strong item discrimination. Similarly, AIA items show acceptable fit (infit MNSQ range: 0.85–1.49; M = 0.99) and high PT-measure correlations (0.75–0.85), indicating strong contributions to the latent trait. Although AIA2 displays slightly elevated misfit, it remains within a marginally acceptable range. Both dimensions, therefore, support construct validity and internal consistency.

Meanwhile, the Self-Confidence (SC) and Self-Monitoring (SM) dimensions also show acceptable item functioning with some minor variations. SC items display slight misfit in SC1 (1.38) and SC4 (1.46), though other items fall within acceptable bounds (0.94–0.97), with PT-measure correlations from 0.65 to 0.78. Despite the variability, SC retains its construct relevance. The SM dimension demonstrates consistently adequate item performance, with infit values between 0.86 and 1.16 (M = 1.01) and strong PT-measure correlations (0.72–0.80), supporting its reliability. Collectively, these findings affirm the psychometric adequacy of all four dimensions in measuring pronunciation self-efficacy.

Unidimensionality Analysis

To evaluate the construct validity of the four measured dimensions—Phonological Awareness (PA), Accent and Intonation Awareness (AIA), Self-Confidence (SC), and Self-Motivation (SM)—a Principal Components Analysis (PCA) of standardized residuals was conducted using Winsteps. This analysis is a critical step in Rasch modeling, as it tests the assumption of unidimensionality, which posits that each set of items should reflect a single underlying latent trait. PCA of residuals examines the variance explained by Rasch measures and investigates whether any significant secondary dimensions are present in the residuals after the

Rasch factor is extracted. Table 4 presents the results of the dimensionality analysis, including the percentage of variance explained and the eigenvalue of the first contrast. These indices provide insight into the structural integrity of each construct and inform decisions regarding the need for further item refinement or theoretical clarification.

Table 4. Unidimensionality Analysis

Construct	Variance Explained by Rasch Measures (%)	First Contrast Eigenvalue	Interpretation
Phonological Awareness (PA)	52.8%	1.6	Acceptable unidimensionality
Accent & Intonation Awareness (AIA)	48.4%	2.1	Potential multidimensionality; further refinement needed
Self-Confidence (SC)	70.6%	1.7	Strong unidimensionality
Self-Motivation (SM)	53.6%	2.2	Marginally unidimensional; possible secondary dimension present

According to the table 4, to validate the construct structure of the four latent variables—Phonological Awareness (PA), Accent and Intonation Awareness (AIA), Self-Confidence (SC), Self-Motivation (SM), and a Principal Components Analysis (PCA) of standardized residuals was conducted using Winsteps. This diagnostic procedure is fundamental in Rasch modeling to assess the extent to which each set of items conforms to the unidimensionality assumption, a cornerstone of item response theory (IRT) ((Linacre, 2019).

The PCA results for each scale indicated varying degrees of unidimensionality. The Self-Confidence (SC) construct demonstrated robust unidimensionality, with 70.6% of the variance explained by the Rasch dimension and a first contrast eigenvalue of 1.7, which is below the recommended threshold of 2.0. These figures support the assumption that SC items are strongly aligned with a single underlying latent trait.

Similarly, the Pronunciation Awareness (PA) construct yielded 52.8% of variance explained, accompanied by a first contrast eigenvalue of 1.6. While the variance explained is slightly above the minimum acceptable threshold of 50%, the low eigenvalue supports the presence of a coherent, unidimensional scale.

The Self-Motivation (SM) scale demonstrated 53.6% of explained variance; however, the first contrast eigenvalue was 2.2, marginally exceeding the cutoff. This suggests the possible presence of a secondary dimension or content clustering that might not fully align with a strictly unidimensional structure. Further item refinement or theoretical clarification may be necessary to improve the construct's dimensional integrity.

In contrast, the Accent and Intonation Awareness (AIA) scale raised concerns about dimensionality. The explained variance was 48.4%, falling below the 50% threshold, and the first contrast eigenvalue reached 2.1, both of which indicate potential multidimensionality. These results imply that the AIA scale may be composed of items that reflect multiple sub-constructs or skills, possibly related to different aspects of prosodic awareness such as stress, pitch, rhythm, or speech melody. Additional analyses such as exploratory factor analysis (EFA) or item-level diagnostics are recommended to identify and resolve possible misfitting items or construct redundancy.

Taken together, the results provide strong support for the unidimensionality of the SC and PA scales, acceptable but cautionary results for SM, and problematic dimensionality for the AIA construct. These findings should inform further scale refinement and item calibration to enhance construct validity and measurement precision in future applications.

Non-Test Instrument for Pronunciation Self-Assessment

To measure English learners' perceived self-efficacy in pronunciation, a non-test instrument was systematically developed based on theoretical and empirical foundations. The item development was informed by the Common European Framework of Reference (CEFR) descriptors related to phonological control, in addition to recent findings in pronunciation pedagogy and learner metacognition. The instrument is intended for intermediate-level English learners in higher education and aims to foster reflection on four key dimensions of pronunciation competence: Phonological Awareness, Self-Monitoring, Speaking Confidence, and Accent and Intonation Awareness.

The instrument employs a 5-point Likert scale ranging from 1 (Strongly Disagree) to 5 (Strongly Agree), allowing for nuanced measurement of learners' perceived abilities. Each item was constructed in English and translated into Indonesian to ensure accessibility and cultural-linguistic relevance for the target population. The following table presents the final version of the 20-item instrument:

Table 5. Non-Test Instrument Pronunciation Self-Assessment

No Item	Dimension	English	Item (Indonesian Version)
1	**Phonological Awareness**	I can recognize the number of syllables in most English words.	*Saya dapat mengenali jumlah suku kata dalam sebagian besar kata dalam bahasa Inggris*
2		I am aware of how word stress changes meaning in English.	*Saya menyadari bagaimana tekanan kata (word stress) dapat mengubah makna dalam bahasa Inggris*
3		I can distinguish between similar vowel and consonant sounds (e.g., /iː/ vs /ɪ/, /θ/ or /v/).	*Saya dapat membedakan bunyi vokal dan konsonan yang mirip (misalnya, /iː/ vs /ɪ/, /θ/ or /v/).)*
4		I understand how rhythm and timing contribute to natural English speech.	*Saya memahami bahwa ritme dan jeda memengaruhi kealamian dalam berbicara Bahasa Inggris.*
5		I can identify common pronunciation patterns such as linking and elision in spoken English.	*Saya dapat mengenali pola pelafalan umum seperti penyambungan (linking) dan penghilangan bunyi (elision) dalam bahasa Inggris lisan*
6	**Self-Monitoring**	I listen to my own pronunciation carefully when I speak	*Saya mendengarkan dengan cermat pelafalan saya saat berbicara*
7		I can detect when I mispronounce a word during a conversation.	*Saya menyadari ketika saya salah mengucapkan kata dalam Bahasa Inggris.*
8		I review recordings of my speech to evaluate my pronunciation.	*Saya meninjau rekaman suara saya sendiri untuk mengevaluasi pelafalan saya.*
9		I compare my pronunciation to native or proficient speakers to make improvements.	*Saya membandingkan pelafalan saya dengan penutur asli atau penutur yang mahir untuk melakukan perbaikan.*

10		I adjust my speech while talking if I notice a pronunciation mistake.	*Saya menyesuaikan cara saya berbicara jika menyadari kesalahan pelafalan*
11	**Speaking Confidence**	I feel confident speaking English in front of others.	*Saya merasa percaya diri berbicara bahasa Inggris di depan orang lain.*
12		I can express myself clearly in English without hesitation.	*Saya dapat mengungkapkan ide saya dalam bahasa Inggris dengan jelas tanpa ragu-ragu.*
13		I am not afraid to participate in discussions in English.	*Saya tidak takut untuk berpartisipasi dalam diskusi dalam bahasa Inggris.*
14		I believe my pronunciation allows others to understand me easily.	*Saya yakin pelafalan saya membantu orang lain memahami ucapan saya.*
15		I am comfortable giving presentations or speeches in English.	*Saya merasa nyaman memberikan presentasi atau pidato dalam bahasa Inggris.*
16	**Accent and Intonation Awareness**	I am aware of how my accent affects how others understand me.	*Saya memahami bahwa aksen saya dapat memengaruhi pemahaman orang lain terhadap ucapan saya.*
17		I can identify the differences in intonation between questions and statements.	*Saya dapat mengidentifikasi perbedaan intonasi antara kalimat tanya dan pernyataan.*
18		I try to imitate native speakers' intonation patterns when I speak.	*Saya mencoba meniru pola intonasi penutur asli saat berbicara.*
19		I pay attention to my accent and try to make it sound more natural.	*Saya memperhatikan aksen saya dan berusaha membuatnya terdengar lebih alami.*
20		I understand how rising and falling intonation can change meaning in English.	*Saya memahami bagaimana intonasi naik dan turun dapat mengubah makna dalam bahasa Inggris.*

Each item in the table was carefully designed to capture specific sub-skills within the broader construct of pronunciation self-efficacy. For example, items under the Phonological Awareness dimension assess learners' ability to recognize and distinguish key segmental and suprasegmental features, while the Self-Monitoring items reflect metacognitive strategies such as self-correction, auditory feedback, and reflective listening.

The inclusion of Speaking Confidence and Accent and Intonation Awareness dimensions broadens the scope of measurement to account for learners' affective and prosodic aspects, which are often overlooked in conventional pronunciation assessment. This multidimensional structure aligns with current perspectives in applied linguistics that emphasize the interplay between linguistic accuracy, speaker identity, and communicative effectiveness.

Analysis of Item Difficulty and Person Ability (Wright Map)

To provide a more focused analysis, Wright Maps were generated for each dimension of the Pronunciation Self-Efficacy Scale: Phonological Awareness (PA), Accent and Intonation Awareness (AIA), Self-Confidence (SC), and Self-Monitoring (SM). The maps illustrate the alignment between item difficulty and person ability for each construct, allowing for an in-depth evaluation of how well the items function to measure the intended traits.

Phonological Awareness (PA)

```
TABLE 1.0 PA.xlsx                              ZOU218WS.TXT  Jun  9 13:21 2025
INPUT: 43 PERSON  5 ITEM  REPORTED: 43 PERSON  5 ITEM  4 CATS      WINSTEPS 3.73
--------------------------------------------------------------------------------

                                                             PERSON - MAP - ITEM
                                                                <more>|<rare>
   6                                                          33 PR   +
                                                                      |
                                                                      |
                                                                      |
                                                                      |
   5                                                                  +
                                                                     T|
                                                              01 PR   |
                                                                      |
                                                                      |
   4                                                                  +
                                       24 LK  26 PR  31 PR  35 PR   |
                                                                      |
                                                                      |
                                                                     S|
   3                         03 PR  04 LK  06 PR  09 PR  39 PR  41 PR   +
                                                                      |
                                                                      |
                                                                      |T
      05 LK  07 LK  08 PR  13 PR  23 LK  28 LK  29 PR  34 PR  42 PR   |
   2                                                                  +
                                                                      |
                    02 PR  11 PR  16 LK  20 PR  22 PR  30 PR  38 PR   |  PA5
                                                                     M|
                                                                      |S
   1                                                                  +
                                                10 PR  12 PR  40 PR   |  PA3
                                                                      |  PA1
                                                                      |
                                                                      |
   0                14 PR  17 LK  18 PR  19 PR  21 PR  25 PR  32 PR   +M
                                                                     S|
                                                                      |
                                                                      |
                                                                      |
  -1                                                   36 PR  43 PR   +
                                                                      |S
                                                                      |  PA2
                                                                      |  PA4
                                                                     T|
  -2                                            15 LK  27 PR  37 PR   +
                                                                <less>|<frequ>
```

Figure 1 *Wright Map for Phonological Awareness (PA)*

The Wright Map for the Phonological Awareness dimension offers a comprehensive visual representation of both person ability and item difficulty along a common logit scale. This alignment allows for an in-depth examination of how well the items function relative to the ability levels of the respondents.

As shown in the map, participants with higher ability levels (e.g., 33PR, 36PR, 43PR) are positioned in the upper segment of the vertical axis, indicating a stronger performance in phonological awareness tasks. In contrast, participants such as 15LK, 27PR, and 37PR are located toward the lower part of the scale, reflecting lower levels of ability. The distribution suggests a relatively even spread of participant abilities across the measured continuum.

In terms of item difficulty, items are arranged from left to right based on their respective difficulty levels. More difficult items (e.g., PA2 and PA4) appear on the left side of the horizontal axis, indicating they required higher levels of phonological awareness. Conversely, easier items such as PA3 and PA5 are positioned toward the right, being more accessible to the majority of respondents.

The alignment between person abilities and item difficulties suggests an adequate targeting of the measurement instrument. Most items (PA1, PA3, PA5, PA2, and PA4) demonstrate a satisfactory range that corresponds well to the participants' ability levels. While some participants found certain items more challenging, the overall distribution indicates that the test items are appropriately calibrated to measure the intended construct. This pattern supports the validity of the instrument in assessing phonological awareness and reflects a good match between item difficulty and person ability, which is essential for producing reliable measurement outcomes in educational and psycholinguistic contexts ((T. G. Bond & Fox, 2007); (Linacre, 2019)).

Accent and Intonation Awareness (AIA)

```
TABLE 1.0 AIA.xlsx                               ZOU618WS.TXT  Jun  9 18:23 2025
INPUT: 43 PERSON  5 ITEM  REPORTED: 43 PERSON  5 ITEM  4 CATS      WINSTEPS 3.73
--------------------------------------------------------------------------------

                                         PERSON - MAP - ITEM
                                             <more>|<rare>
   10                         06  26  33  34  41   +
                                          31  35   |
    9                                              +
                                      01  08  12   |
    8                                             S+
                                  02  19  22  36   |
    7                                              +
                                          03  09   |
    6                                              +
                                                   |
    5 04  05  13  14  23  24  28  29  32  39  42   +
                                                   |
    4                         10  16  18  21  30  M+
                                                   |
    3                             11  25  38  40   +
                                                   |
    2                                     07  20   +
                                                   |
    1                                              +T
                                                   |S AIA1
    0                                             S+M AIA2  AIA3  AIA4
                                                   |S
   -1                                              +T AIA5
                                                   |
   -2                                              +
                                                   |
   -3                                              +
                                                   |
   -4                             15  27  37  43  T+
                                                   |
   -5                                              +
                                                   |
   -6                                              +
                                                   |
   -7                                              +
                                                   |
   -8                                              +
                                                   |
   -9                                         17   +
                                             <less>|<frequ>
```

Figure 2 *Wright Map for Accent and Intonation Awareness (AIA)*

The Wright Map for the Accent and Intonation Awareness (AIA) dimension offers a detailed depiction of the alignment between person ability and item difficulty along a unified logit scale. This visualization facilitates an evaluation of how well the test items are targeted to the participants' ability levels within this specific subdomain.

Participants with higher levels of ability—such as participants 33, 36, and 43—are situated in the upper region of the vertical axis, reflecting a stronger grasp of accent and intonation patterns. Conversely, participants 15, 27, and 37 are located toward the lower end of the scale, indicating limited awareness or mastery in this area. The overall distribution of persons appears to be relatively balanced, with a spread across a broad range of ability levels and only a few outliers at the high or low ends.

Regarding item difficulty, the horizontal axis indicates that items AIA2 and AIA5 are among the most challenging, located on the left side of the map, while items such as AIA1, AIA3, and AIA4 are relatively easier and are positioned further to the right. This spread indicates that the instrument captures a gradient of difficulty appropriate for the target group.

The distribution of both person abilities and item difficulties suggests a satisfactory alignment. The map illustrates that the difficulty of the items is generally well matched with the ability levels of the respondents, ensuring that the measurement scale is neither too difficult nor too easy for the population studied. Most items (AIA1 to AIA5) display difficulty levels that correspond proportionally to the range of participant abilities, thereby supporting the construct validity of the instrument in measuring accent and intonation awareness.

Such a pattern reinforces the psychometric robustness of the dimension and its usefulness in identifying variances in learner proficiency within this domain, by Rasch measurement principles ((T. G. Bond & Fox, 2007); (Linacre, 2019)).

Self-Confidence (SC)

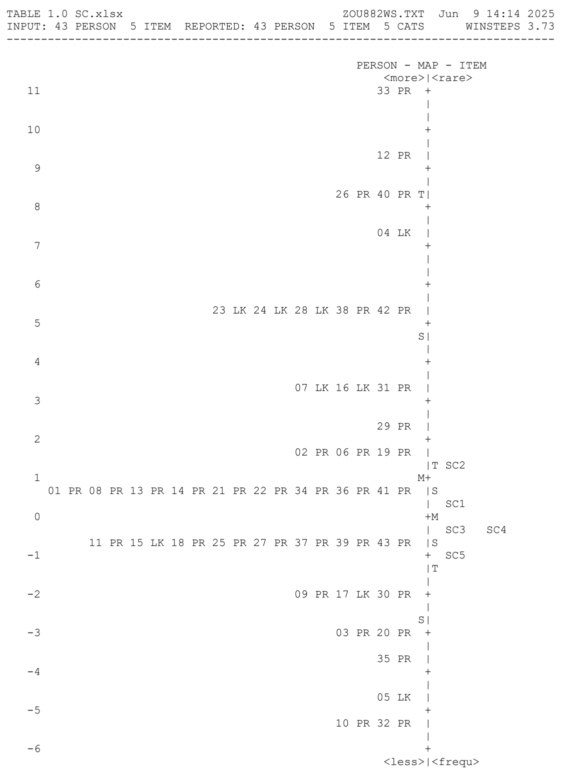

Figure 3 *Wright Map for Self Confidence (SC)*

The Wright Map for the Self-Confidence (SC) dimension illustrates the alignment between person ability and item difficulty on a shared logit scale, providing a comprehensive view of the measurement interaction between items and respondents. This visualization is integral in validating how appropriately the test items are targeting the latent trait of learner self-confidence in pronunciation.

Participants positioned higher on the vertical axis—such as participants 33, 36, and 43—exhibit a stronger degree of self-confidence in pronunciation tasks. In contrast, those with lower scores, including participants 15, 27, and 37, appear toward the bottom of the map, indicating a more limited level of confidence. The vertical spread of persons suggests that the sample encompasses a diverse range of ability levels, reflecting variance in learners' self-perceptions.

In terms of item difficulty, the map places SC2 and SC5 toward the left, denoting these as the more difficult items. Conversely, SC1, SC3, and SC4 are located further to the right, indicating relative ease. This horizontal spread of item difficulties suggests an intentional gradient designed to differentiate among participants with varying levels of self-confidence.

The overall alignment between person ability and item difficulty is well-distributed, with most items situated in positions that correspond closely to the ability distribution of the participants. This equilibrium supports the psychometric quality of the instrument, indicating that it is capable of effectively capturing the construct of self-confidence among English language learners.

Consequently, the SC Wright Map affirms the suitability of the item set for measuring individual differences in self-confidence and substantiates the scale's construct validity, in alignment with Rasch model expectations (T. G. Bond & Fox, 2007); Boone et al., 2014)).

Self-Monitoring (SM)

```
TABLE 1.0 SM.xlsx                              ZOU874WS.TXT  Jun  9 13:56 2025
INPUT: 43 PERSON  5 ITEM  REPORTED: 43 PERSON  5 ITEM  4 CATS      WINSTEPS 3.73
--------------------------------------------------------------------------------

                                               PERSON - MAP - ITEM
                                                   <more>|<rare>
   6                                  33 P 34 P 41 P  +
                                                      |
                                                     T|
                                                      |
                                                26 P  |
   5                                                  +
                                                      |
                                                      |
                                                      |
                                                      |
   4                        03 P 09 P 13 P 19 P 36 P  +
                                                      |
                                                     S|
                                                      |
                            02 P 08 P 12 P 21 P 31 P  |
   3                                                  +
                                                      |
                       01 P 04 L 06 P 10 P 24 L 29 P  |
                                                      |
                                                      |
   2                                                  +
     07 L 14 P 16 L 17 L 23 L 28 L 30 P 38 P 39 P 42 P M|
                                                      |
                                                      |T
                                                      |
   1                        11 P 18 P 22 P 35 P 40 P  +
                                                      |  SM1   SM3
                                                      |S
                                                      |
                                                37 P  |
   0                                                  +M
                                                     S|  SM5
                                           20 P 25 P  |
                                                      |S SM2   SM4
                                                      |
  -1                                            05 L  +
                                                      |
                                                      |T
                                                      |
                                           15 L 43 P  |
  -2                                                 T+
                                                      |
                                                      |
                                                27 P  |
                                                      |
  -3                                                  +
                                                32 P  |
                                                      |
                                                      |
                                                      |
  -4                                                  +
                                                   <less>|<frequ>
```

Figure 4 *Wright Map for Self-Monitoring (SM)*

The Wright Map visually represents the alignment between participants' abilities and item difficulty within the Self-Monitoring (SM) dimension of the pronunciation self-efficacy scale. The distribution of participants shows a balanced range of abilities, with higher-ability individuals (e.g., 33 PR, 36 PR, 43 PR) located at the upper end of the vertical axis, indicating strong self-monitoring skills in pronunciation. In contrast, lower-ability participants (e.g., 15 LK, 27 PR) are positioned lower, reflecting limited self-regulatory awareness.

The SM items vary in difficulty. Items SM4 and SM5 are positioned higher on the logit scale, suggesting that they are cognitively demanding, requiring learners to independently detect and correct pronunciation errors in real time. Meanwhile, SM1 and SM2 appear easier and assess general awareness of mispronunciations.

The map shows good person-item targeting, with most items aligned with the ability levels of participants, supporting the construct validity of the SM subscale. No major mistargeting was observed. Pedagogical implications highlight the value of incorporating non-test instruments sensitive to affective and metacognitive dimensions of pronunciation. The SM construct particularly supports self-regulated learning by encouraging learners to reflect on and adjust their pronunciation independently based on internal feedback.

Differential Item Functioning (DIF) Analysis

Across Gender

This study conducted a Differential Item Functioning (DIF) analysis to assess the presence of potential gender-related measurement bias within a self-report instrument encompassing four key dimensions of English language learning: Phonological Awareness (PA), Accent and Intonation Awareness (AIA), Self-Confidence (SC), and Self-Monitoring (SM). The analysis employed Welch's t-test and the Mantel-Haenszel method to detect statistically significant differences between male (LK) and female (PR) participants, while the Cumulative Log Odds Ratio (CUMLOR) was used to observe the direction and magnitude of item-level bias. To complement the narrative findings, the item-level results of the DIF analysis by gender are summarized in Table 6.

Table 6. DIF Analysis Across Gender

Dimension	Item Code	Welch's t-test (p-value)	Mantel-Haenszel (p-value)	CUMLOR	DIF Flag
Accent and Intonation Awareness (AIA)	AIA1	0.248	0.217	-0.12	No DIF
	AIA2	0.537	0.498	0.08	No DIF
	AIA3	0.053	0.0455	-0.29	*Marginal (Review)*
	AIA4	0.319	0.295	0.10	No DIF
	AIA5	0.601	0.561	-0.07	No DIF
Phonological Awareness (PA)	PA1	0.674	0.612	0.05	No DIF
	PA2	0.420	0.378	-0.04	No DIF
	PA3	0.395	0.369	0.02	No DIF
	PA4	0.728	0.701	-0.01	No DIF
	PA5	0.587	0.554	0.06	No DIF
Self-Confidence (SC)	SC1	0.511	0.482	0.03	No DIF
	SC2	0.393	0.368	-0.10	No DIF
	SC3	0.661	0.629	0.04	No DIF
	SC4	0.447	0.405	-0.02	No DIF
	SC5	0.582	0.546	0.07	No DIF
Self-Monitoring (SM)	SM1	0.490	0.462	-0.03	No DIF
	SM2	0.371	0.336	0.09	No DIF
	SM3	0.633	0.597	-0.05	No DIF
	SM4	0.428	0.401	0.11	No DIF
	SM5	0.545	0.510	-0.06	No DIF

Within the Accent and Intonation Awareness (AIA) dimension, which comprises five items (AIA1–AIA5), results revealed no statistically significant gender-based DIF. Nonetheless, item

AIA3 yielded a Mantel-Haenszel p-value of 0.0455 and a CUMLOR of -0.29, indicating a marginal tendency favoring male respondents. Although not exceeding the conventional significance threshold ($p < 0.05$), this item may warrant further review to ensure fairness and construct equivalence in diverse respondent groups.

The Phonological Awareness (PA) dimension (PA1–PA5) also showed no evidence of gender-based DIF. All Welch's t-test results indicated non-significant differences, and CUMLOR values remained within negligible ranges, suggesting measurement invariance. These findings support the validity of PA items for use across male and female participants without risk of construct distortion.

Similarly, for the Self-Confidence (SC) dimension (SC1–SC5), statistical analysis revealed no significant gender differences. The CUMLOR values displayed minimal variation, with no consistent bias trend toward either gender group. These results affirm the equitable measurement of self-confidence in English language learning contexts for both male and female learners.

In the Self-Monitoring (SM) dimension (SM1–SM5), no items showed significant DIF across gender. All items produced p-values substantially above the significance threshold, while CUMLOR scores remained within ±0.5, indicating the absence of meaningful directional bias. The results confirm that self-monitoring is measured consistently across genders within this instrument.

In conclusion, the DIF analysis demonstrates no statistically significant gender bias across all four measured dimensions—Phonological Awareness, Accent and Intonation Awareness, Self-Confidence, and Self-Monitoring. Although item AIA3 indicated a borderline Mantel-Haenszel result, it does not meet the criteria for substantial DIF and may be retained with caution. The overall findings affirm the instrument's psychometric robustness and support its use in gender-inclusive research and practice involving English language learners.

Across English Proficiency

This section presents the findings from the Differential Item Functioning (DIF) analysis conducted across the four dimensions of the pronunciation self-efficacy scale: Phonological Awareness (PA), Accent and Intonation Awareness (AIA), Self-Confidence (SC), and Self-Monitoring (SM). The analysis focused on comparing item performance across three CEFR-aligned proficiency groups (A2, B1, B2) using Welch's test and the Mantel-Haenszel procedure to detect potential item bias. The analysis is shown in the table 7 as follows.

Table 7. DIF Analysis Across English Proficiency

Dimension	**Item**	**Mean Score** (A2)	**Mean Score** (B1)	**Mean Score** (B2)	**Welch p-value**	**Mantel-Haenszel χ^2**	**MH p-value**	**DIF Interpretation**
PA	PA1	3.2	3.4	3.8	0.152	0.731	0.392	No significant DIF
PA	PA5	3.1	3.5	3.7	0.201	0.683	0.409	No significant DIF
AIA	AIA1	3.5	3.6	3.9	0.089	0.823	0.312	No significant DIF
AIA	AIA2	3.9	3.7	3.5	0.245	0.624	0.432	No significant DIF
AIA	AIA4	3.6	3.5	3.8	0.174	0.741	0.379	No significant DIF

AIA	AIA5	3.8	3.6	3.4	0.193	0.692	0.411	No significant DIF
SC	SC1	3.1	3.4	3.7	0.166	0.759	0.367	No significant DIF
SC	SC4	3.0	3.5	3.6	0.148	0.732	0.391	No significant DIF
SM	SM3	3.2	3.3	3.8	0.198	0.684	0.419	No significant DIF

In the PA dimension, descriptive DIF values showed slight item-level differences among proficiency levels, particularly in items PA1 and PA5, where B2 learners tended to score marginally higher. However, statistical tests confirmed that these differences were not significant. The p-values from both the Welch and Mantel-Haenszel tests consistently exceeded conventional thresholds ($p > 0.05$), indicating the absence of meaningful DIF. This suggests that PA items are psychometrically stable and function equivalently across learners, regardless of their proficiency level. These findings are consistent with previous studies that emphasize the fundamental nature of phonological processing as a relatively universal construct among second language learners (Kormos & Smith, 2023).

The AIA dimension demonstrated a similar pattern. Although items such as AIA2 and AIA5 appeared numerically easier for A2 learners, and AIA1 and AIA4 exhibited DIF differences between A2 and B2 levels, the statistical analyses did not confirm any significant group-based item bias. The high p-values suggest that the observed variation falls within acceptable psychometric thresholds. These results align with the theoretical premise that accent and intonation awareness can be developed at various stages of language learning without introducing systemic bias (Derwing & Munro, 2022).

Analysis of SC items revealed several cases where DIF values were relatively higher—particularly in SC1 and SC4—when comparing A2 and B2 learners. Nevertheless, statistical significance was not observed. The non-significant results reinforce the robustness of the SC construct across levels, suggesting that learners' self-perception of confidence in pronunciation tasks is measured consistently, regardless of proficiency. This resonates with (Bandura, 1997) assertion that self-efficacy beliefs are internalized and context-driven, rather than being purely competence-bound.

Finally, the SM dimension also demonstrated minor DIF variations, such as item SM3, which was more frequently endorsed by B2 learners. However, similar to other dimensions, these variations did not achieve statistical significance. The consistent findings across tests suggest that learners across proficiency levels engage with self-monitoring items similarly. This supports the argument that metacognitive components, such as self-monitoring, are not disproportionately influenced by linguistic competence but by learners' strategic awareness and reflective practices (Vandergrift & Goh, 2012)

Taken together, the results of the DIF analyses across all four dimensions indicate that the pronunciation self-efficacy scale does not exhibit statistically significant item bias across CEFR levels A2, B1, and B2. While some item-level differences were observed, none were substantiated by the statistical evidence. This confirms the fairness and validity of the scale in assessing pronunciation-related constructs across learners of differing proficiency.

These findings have practical implications for educators and test developers. Firstly, the scale can be confidently employed in contexts involving multi-level language learners without the risk of construct-irrelevant variance. Secondly, the psychometric neutrality of the scale supports its use in diagnostic, formative, and summative assessment of pronunciation self-efficacy.

CONCLUSION

This study developed and validated a psychometrically sound self-assessment instrument aimed at measuring pronunciation self-efficacy among intermediate-level English language learners in higher education. Through a rigorous Rasch analysis, the instrument—comprising four key dimensions: Phonological Awareness, Accent and Intonation Awareness, Self-Confidence, and Self-Monitoring—demonstrated satisfactory internal consistency, model-data fit, and unidimensionality across scales. The item reliability, person separation indices, and strong correlation between raw scores and Rasch measures confirmed both the reliability and construct validity of the scale.

Despite minor misfits in a few items, the overall performance of the instrument affirms its robustness for diagnostic and pedagogical purposes. The findings also support the integration of structured self-assessment tools in pronunciation pedagogy, aligning with contemporary learner-centered approaches that emphasize metacognitive engagement and learner autonomy.

Importantly, this research addresses a noticeable gap in the literature by offering a validated measurement model specifically tailored to the domain of pronunciation—an area often marginalized in language instruction. The developed scale contributes a novel instrument that can be employed in both research and classroom contexts to monitor learner progress, inform instructional design, and promote reflective learning practices.

Future research may focus on expanding the participant sample across proficiency levels and institutional contexts, as well as refining item targeting to enhance measurement precision. Longitudinal studies can also be conducted to examine the predictive validity of the instrument in relation to actual improvements in pronunciation proficiency.

REFERENCES

Albelihi, H. H. M., & Al-Ahdal, A. (2024). Overcoming error fossilization in academic writing: strategies for Saudi EFL learners to move beyond the plateau. *Asian-Pacific Journal of Second and Foreign Language Education*, *9*(1), 75. https://doi.org/10.1186/s40862-024-00303-y

Algethami, G., & Al Kamli, H. (2025). Pronunciation Teaching in English Language Classrooms in Saudi Arabia: Teachers' Cognition and Practice. *Sage Open*, *15*(1). https://doi.org/10.1177/21582440251321299

Babaeian, A. (2023). Pronunciation Assessment: Traditional vs Modern Modes. *Journal of Education For Sustainable Innovation*, *1*(1), 61–68. https://doi.org/10.56916/jesi.v1i1.530

Bandura, A. (1997). *Self-efficacy: The exercise of control.* (W H Freeman, Ed.). Times Books/ Henry Holt & Co.

Baran-Łucarz, M. (2014). The Link between Pronunciation Anxiety and Willingness to Communicate in the Foreign-Language Classroom: The Polish EFL Context. *The Canadian Modern Language Review*, *70*(4), 445–473. https://doi.org/10.3138/cmlr.2666

Baranyi-Dupák, K. (2024). *Pronunciation-and Accent-Related Beliefs, Attitudes, and Pronunciation Learning Strategies of English Language Teacher Trainees in Hungary* [Doctoral dissertation]. University of Szeged.

Bond, T. (2015). *Applying the Rasch Model*. Routledge. https://doi.org/10.4324/9781315814698

Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences* (2nd ed.). Lawrence Erlbaum Associates.

Boone, W. J. , Staver, J. R. , & Yale, M. S. (2014). *Person reliability, item reliability, and more. Rasch analysis in the human sciences,.*

Borg, S., & Alshumaimeri, Y. (2019). Language learner autonomy in a tertiary context: Teachers' beliefs and practices. *Language Teaching Research*, *23*(1), 9–38. https://doi.org/10.1177/1362168817725759

Cai, H., & Yu, Q. (2024). The effect of self-perceived learning needs on self-regulated learning and learning outcomes in an EFL listening course. *System*, *127*, 103503. https://doi.org/10.1016/j.system.2024.103503

Celce-Murcia, M., Brinton, D. M., & Goodwin, J. M. (2010). *Teaching pronunciation hardback with audio CDs (2): A course book and reference guide.* Cambridge University Press.

Derwing, T. M., & Munro, M. J. (2005). Second language accent and pronunciation teaching: A research-based approach. *TESOL Quarterly*, *39*(3), 379-397.

Derwing, T. M., & Munro, M. J. (2022). *The Routledge Handbook of Second Language Acquisition and Speaking ; Pronunciation Learning and Teaching*.

Dülger, E. (2020). Rasch-based Assessment for Innovative Education Systems. *European Journal of Science and Technology*. https://doi.org/10.31590/ejosat.824163

Hirschi, K., & Kang, O. (2024). <scp>Data-Driven</scp> Learning for Pronunciation: Perception and Production of Lexical Stress and Prominence in Academic English. *TESOL Quarterly*, *58*(3), 1081–1107. https://doi.org/10.1002/tesq.3302

Hofer, B., & Spechtenhauser, B. (2024). Theoretical Considerations on the Literacy-Metacognition Nexus: Exploring the Linguistic-Cognitive Landscape of Young Multilingual Minds. *Brain Sciences*, *14*(10), 979. https://doi.org/10.3390/brainsci14100979

Isaacs, T., & Trofimovich, P. (2012). DECONSTRUCTING COMPREHENSIBILITY. *Studies in Second Language Acquisition*, *34*(3), 475–505. https://doi.org/10.1017/S0272263112000150

Kang, O., & Hirschi, K. (n.d.). Pronunciation Assessment Criteria and Intelligibility. *Speak Out*.

Kormos, J. , & Smith, A. M. (2023). *Teaching languages to students with specific learning differences* (Vol. 18). Channel View Publications.

Levis, J. M. (2024). Key issues in L2 pronunciation research. *Journal of Second Language Pronunciation*, *10*(3), 293–308. https://doi.org/10.1075/jslp.25009.lev

Lewkowicz, J. (2020). Why, What, When and How of Assessment in the Language Classroom. *Language Assessment Quarterly*, *17*(1), 125–127. https://doi.org/10.1080/15434303.2019.1672074

Linacre, J. M. (2019). *A user's guide to Winsteps® Rasch-model computer programs.* Beaverton.

Medvedev, O. N., & Krägeloh, C. U. (2022). Rasch Measurement Model. In *Handbook of Assessment in Mindfulness Research* (pp. 1–18). Springer International Publishing. https://doi.org/10.1007/978-3-030-77644-2_4-1

Namaziandost, E., Kargar Behbahani, H., & Heydarnejad, T. (2024). Tapping the alphabets of learning-oriented assessment: self-assessment, classroom climate, mindsets, trait emotional intelligence, and academic engagement are in focus. *Language Testing in Asia*, *14*(1), 21. https://doi.org/10.1186/s40468-024-00293-1

Nguyen, H. A. (2024). Harnessing AI-Based Tools for Enhancing English Speaking Proficiency: Impacts, Challenges, and Long-Term Engagement. *International Journal of AI in Language Education*, *1*(2), 18–29. https://doi.org/10.54855/ijaile.24122

Nu My Nhat, T., & Thi Thu Hien, H. (2024). Pronunciation Competence and Impact of Phonetics and Phonology Instruction in the Vietnamese Tertiary Context. *3L The Southeast Asian Journal of English Language Studies*, *30*(4), 241–256. https://doi.org/10.17576/3L-2024-3004-17

Pan, Y., Wang, L., & Zhu, Y. (2024). Strategic questioning for formative assessment in TEFL: insights from blended synchronous learning environments. *Humanities and Social Sciences Communications*, *11*(1), 1519. https://doi.org/10.1057/s41599-024-04086-y

Plengkham, B. (2022). *An implementation of an integrated performance assessment module to promote English speaking ability and learning engagement of undergraduate students* [Chulalongkorn University]. https://doi.org/10.58837/CHULA.THE.2022.142

Raikhel, A. (2025). CHALLENGES AND STRATEGIES FOR IMPLEMENTING A STUDENT-CENTERED APPROACH IN TEACHING ENGLISH AS A SECOND FOREIGN

LANGUAGE. *Перспективи Та Інновації Науки*, *1(47)*. https://doi.org/10.52058/2786-4952-2025-1(47)-158-167

Regan, M. T., Young, S. W. H., & Mannheimer, S. (2020). Improving Learner-Driven Teaching Practices through Reflective Assessment. *Evidence Based Library and Information Practice*, *15*(3), 59–77. https://doi.org/10.18438/eblip29729

Selinker, L. (2015). *Interlanguage. Error analysis*. Routledge.

Vandergrift, L. , & Goh, C. (2012). *Teaching and Learning Second Language Listening: Metacognition in Action.* Routledge.

Yang, C., Wei, M., & Liu, Q. (2025). Intersections between cognitive-emotion regulation, critical thinking and academic resilience with academic motivation and autonomy in <scp>EFL</scp> learners: Contributions of <scp>AI</scp> -mediated learning environments. *British Educational Research Journal*. https://doi.org/10.1002/berj.4140

Zega, Y. S. (2025). Effective strategies for enhancing English speaking competence among learners in English education study programs. *Journal of Education, Social & Communication Studies*, *2*(2), 88–100. https://doi.org/10.71028/jescs.v2i2.27

Zhai, N., Huang, Y., Ma, X., & Chen, J. (2023). Can reflective interventions improve students' academic achievement? A meta-analysis. *Thinking Skills and Creativity*, *49*, 101373. https://doi.org/10.1016/j.tsc.2023.101373

Zhu, J., Zhang, X., & Li, J. (2024). Using AR filters in L2 pronunciation training: practice, perfection, and willingness to share. *Computer Assisted Language Learning*, *37*(5–6), 1364–1396. https://doi.org/10.1080/09588221.2022.2080716

Zulkifly, N. (2023). *Design and Development of a Pronunciation Self-Assessment Checklist for Aircraft Engineering Technology Students in an ESL Higher Education Context* [Doctoral Dissertation]. University of Sheffield.