Paradigm Academic Press
Research and Advances in Education
ISSN 2788-7057
JUN. 2026 VOL.5, NO.2

Are We Really Listening? Development and Psychometric Validation of a Five-Stage Listening Process Scale for Higher Education Students in Region II, Philippines

Mark Kevin Astrero[1], Saranay I. Doyaoen[1], Darius King Galo[1], Aileen Melody De Vera[1] & Titov Dela Cruz[2]

[1] Northeastern College, Santiago City, Isabela, Philippines

[2] University of La Salette Incorporated High School, Santiago City, Isabela, Philippines

Correspondence: Mark Kevin Astrero, Northeastern College, Santiago City, Isabela, Philippines.

doi:10.63593/RAE.2788-7057.2026.06.004

Abstract

Listening is central to classroom learning yet rarely measured with locally validated tools in Philippine higher education. This study developed and validated a 50-item Listening Process Scale (LPS), grounded in Brownell's stage model (receiving, understanding, remembering, evaluating, responding), using 314 students from higher education institutions in Region II, Philippines. Sampling adequacy was excellent (KMO = .968). Exploratory and confirmatory factor analyses converged on a hierarchical structure: five correlated first-order factors loading strongly (.88–.91) onto one General Listening Process factor, rather than five independent stages. Reliability was excellent overall (α = .976) and per subscale (.900–.937), though elevated inter-item correlations suggest some redundancy. Structural equation modeling showed the General factor significantly predicted academic performance (β = .212, $p < .001$). Students self-reported High listening levels across all stages (M = 3.59). The LPS is a reliable, academically relevant measure, best interpreted through its Overall score pending further item refinement.

Keywords: listening process, questionnaire validation, exploratory factor analysis, HURIER model, higher education institutions, Region II Philippines, scale development

1. Introduction

Among the four macro-communication skills — listening, speaking, reading, and writing — listening is simultaneously the most frequently used and the least deliberately taught (Wolvin & Coakley, 1996; Gilakjani & Ahmadi, 2011). College and university students spend a substantial share of their academic day receiving spoken information through lectures, discussions, and peer interaction, yet listening competence is seldom assessed with the same rigor applied to reading comprehension or written expression. This asymmetry is consequential: instructors, program chairs, and student-support units in Philippine higher education institutions (HEIs) have few locally validated instruments with which to diagnose where students' listening difficulties actually lie — at the point of attention, comprehension, retention, evaluation, or response.

The present study addresses this gap by developing and validating a Listening Process Scale (LPS) for HEI students in Region II (Cagayan Valley), Philippines. Rather than treating listening as a single undifferentiated trait, the LPS operationalizes listening as a sequence of five interrelated cognitive-behavioral stages — receiving/attending, understanding, remembering, evaluating, and responding — consistent with stage models of the listening process that have long informed communication pedagogy (Brownell, 1985, 1996). Validating such a stage-based instrument locally serves two purposes. First, it gives Philippine HEIs a diagnostic tool sensitive enough to identify which stage of the listening process students struggle with most, rather than a single global score that obscures where intervention should be targeted. Second, it contributes empirical evidence, from an under-represented Southeast Asian population, to the broader question of whether listening stage models replicate as separable statistical dimensions or collapse into fewer, broader factors when measured with self-report items — a question with direct relevance to how listening curricula and assessments should be structured.

Specifically, this study sought to: (1) examine the factorial structure of the 50-item LPS using exploratory factor analysis; (2) evaluate the internal consistency reliability of the overall scale and its five hypothesized subscales; (3) describe the self-reported level of each listening stage among Region II HEI students; and (4) examine whether LPS scores differ meaningfully across student sex, age bracket, first-language background, and academic performance, as a source of criterion-related validity evidence.

2. Theoretical Framework and Related Literature

2.1 Stage Models of the Listening Process

Brownell's HURIER model remains one of the most widely cited frameworks for conceptualizing listening as a multi-stage process rather than a single skill, decomposing listening into hearing, understanding, remembering, interpreting, evaluating, and responding (Brownell, 1985, 1996). The model treats these stages as interdependent rather than strictly sequential: weaknesses at an early stage (e.g., attentional lapses during hearing) constrain what can be accomplished at later stages (e.g., accurate evaluation or an appropriate response). The present instrument adapts this tradition into five measurable dimensions — collapsing hearing/attending into a single receiving stage and folding interpretive judgments (attending to tone, context, and underlying intent) into the understanding stage — a simplification intended to make the construct tractable for self-report measurement with a student population while preserving the process logic of the original model.

2.2 Measuring Listening: Prior Instruments and Open Questions

Several validated listening instruments exist in the communication literature, including the Watson–Barker Listening Test (Bodie, Worthington, & Fitch-Hauser, 2011), the Listening Styles Profile–16 and its revision (Bodie & Worthington, 2010; Bodie, Worthington, & Gearhart, 2013), and the Active-Empathic Listening Scale (Bodie, 2011). Collectively, this body of work has repeatedly found that self-report listening measures are prone to producing a large, dominant general factor, with narrower conceptually distinct facets emerging only when items are written with sufficient specificity and sample sizes are adequate (Bodie, Worthington, Imhof, & Cooper, 2008; Bodie, 2013). This measurement literature — developed almost entirely on North American samples — has not, to the researchers' knowledge, been extended to a stage-based instrument validated on a Filipino HEI population, which the present study addresses.

2.3 Listening and Academic Performance in the Philippine Context

Locally, listening difficulties have been repeatedly linked to broader academic and language-learning outcomes. Descriptive studies of Philippine secondary and tertiary learners report that listening proficiency lags behind other language skills and correlates with academic indicators such as general weighted average (De Vera et al., 2018, as cited in academic listening research on Philippine learners; Aguacito, Macatabon, & Imbang-Agni, 2025). These studies, however, have generally relied on researcher-made instruments without a formal psychometric validation stage, which limits the confidence with which their findings can be generalized or their instruments reused. The present study responds to this methodological gap by subjecting a stage-based listening instrument to a full sequence of sampling-adequacy testing, exploratory factor analysis, reliability analysis, and known-groups validation.

3. Methods

3.1 Research Design

This study employed a descriptive-validation, cross-sectional survey design, combining psychometric scale-validation procedures with a descriptive assessment of self-reported listening behavior.

3.2 Participants and Sampling

Participants were 314 students drawn from higher education institutions across Region II, Philippines, recruited through an online, self-administered questionnaire distributed via institutional and social-media channels (convenience sampling). Table 1 summarizes the demographic profile of the sample. The sample was predominantly female (62.4%) and in the 19–21 age bracket (58.0%); most respondents identified Filipino/Tagalog as their family's first language (67.5%), and the modal academic-performance bracket was 85–89 (43.6%).

Table 1. Demographic Profile of Respondents

Profile Variable	Category	n	%
Sex	Female	196	62.4
	Male	118	37.6
Age Bracket	16–18	103	32.8
	19–21	182	58.0
	22 and above	29	9.2
Family’s First Language	Filipino/Tagalog	212	67.5
	Iloco	30	9.6
	Visaya	29	9.2
	English	26	8.3
	Other local dialects	17	5.4
Academic Performance (General Average)	below 80	11	3.5
	80–85	71	22.6
	85–89	137	43.6
	90–100	95	30.3

3.3 Instrument

The LPS is a 50-item, five-point Likert scale (1 = strongly disagree/almost never, 5 = strongly agree/almost always) composed of five researcher-defined, 10-item subscales corresponding to the receiving/attending, understanding, remembering, evaluating, and responding stages of listening. Items were written in behaviorally specific, first-person statements (e.g., “I give my full attention to the speaker without being distracted by my phone or other devices”; “I recall the key points of a conversation shortly after it ends”) to reduce the abstraction typical of trait-style listening items.

3.4 Data Quality Screening

Because the questionnaire was self-administered online without proctoring, response patterns were screened for evidence of careless or inattentive responding prior to the factor-analytic and reliability analyses, following recommended practice for online survey data (Meade & Craig, 2012). Twenty-three respondents (7.3% of the sample) provided the identical rating to all 50 items — a classic “straight-lining” pattern that artificially inflates inter-item correlations and can distort factor solutions. These 23 cases were retained for the descriptive-level and reliability analyses (consistent with common practice, since Cronbach’s alpha and mean-level description are comparatively robust to this pattern) but were excluded from the exploratory factor analysis and the Kaiser–Meyer–Olkin/Bartlett sampling-adequacy tests, which are more sensitive to spurious covariance; the resulting factor-analytic sample was N = 291. Both the full-sample and screened-sample results are reported for transparency.

3.5 Data Analysis

Sampling adequacy was evaluated using the Kaiser–Meyer–Olkin (KMO) measure and Bartlett’s test of sphericity. Because parallel analysis has been shown to outperform simpler heuristics such as the Kaiser eigenvalue-greater-than-one rule for determining the number of factors to retain (Watkins, 2018), the number of factors was determined using Horn’s parallel analysis (100 random-data replications) before extracting factors via exploratory factor analysis (EFA) using minimum-residual (minres) estimation with promax (oblique) rotation, appropriate given the theoretically correlated nature of sequential listening stages. Mean inter-item and inter-scale correlations were computed to characterize item redundancy and discriminant separation among subscales (Clark & Watson, 1995).

The measurement structure suggested by the EFA was then formally tested with confirmatory factor analysis (CFA), estimated by maximum likelihood in the Python package semopy (Igolkina & Meshcheryakov, 2020). Four nested models were compared: (1) the a priori five-factor correlated CFA model corresponding to the original subscale design; (2) an EFA-informed five-factor correlated CFA model built from the empirical item groupings; (3) a second-order (hierarchical) CFA model with the five first-order factors loading onto a single General Listening Process factor; and (4) a full structural equation model (SEM) extending the second-order model with a structural regression path from the General factor to students’ academic performance, operationalized as the numeric midpoint of their reported general-average bracket. Model fit was evaluated using the chi-square test, Comparative Fit Index (CFI), Tucker–Lewis Index (TLI), and Root Mean Square Error of Approximation (RMSEA), following conventional joint-criteria guidelines (CFI/TLI ≥ .90 acceptable, ≥ .95 good; RMSEA ≤ .08 acceptable, ≤ .06 good; Hu & Bentler, 1999).

Internal consistency was assessed with Cronbach's alpha for the overall scale and each subscale, interpreted with attention to the caution that very high alpha values can themselves signal item redundancy rather than pure reliability (Taber, 2018). Because Shapiro–Wilk testing indicated departure from normality in the composite scores, group comparisons used the nonparametric Mann–Whitney U test (two-group comparisons: sex) and the Kruskal–Wallis H test (comparisons across three or more groups: age bracket, first language, academic performance). All analyses were conducted in Python 3.12 using pandas, SciPy, factor_analyzer, and semopy.

3.6 Ethical Considerations

Participation was voluntary and anonymous; respondents' email addresses, collected only for survey-platform administration, were removed prior to analysis and are not reported. No personally identifying information is presented in this manuscript.

4. Results

4.1 Sampling Adequacy

The overall Kaiser–Meyer–Olkin measure of sampling adequacy was .968 for the full sample (N = 314) and .963 for the screened sample (N = 291), both well above the conventional "marvelous" threshold of .90 (Watkins, 2018). Bartlett's test of sphericity was significant for both samples (full sample: $\chi^2(1225) = 11{,}411.63$, $p < .001$; screened sample: $\chi^2(1225) = 9{,}718.90$, $p < .001$), confirming that the item correlation matrix was suitable for factor analysis.

4.2 Number of Factors and Factor Structure

Horn's parallel analysis on the screened sample indicated that only the first three empirical eigenvalues exceeded the 95th percentile of eigenvalues generated from random data of the same dimensions (real eigenvalues of 21.75, 2.47, and 1.95 versus random-data thresholds of 1.98, 1.87, and 1.80, respectively), suggesting a strong general listening factor with two additional, narrower factors. The conventional Kaiser criterion (eigenvalue > 1) was considerably more liberal, suggesting seven factors — illustrating the well-documented tendency of that criterion to over-extract (Watkins, 2018).

Given the strength of the theoretical five-stage model and the interpretability criterion recommended alongside statistical stopping rules (Watkins, 2018), a five-factor solution was also examined and is reported as the primary structural result, with the three-factor parallel-analysis solution noted as a more conservative alternative. The five-factor promax solution explained 43.2% of cumulative variance (Factor 1 = 12.6%, Factor 2 = 11.1%, Factor 3 = 8.2%, Factor 4 = 7.3%, Factor 5 = 4.0%) and produced the pattern summarized in Table 2.

Table 2. Summary of the Five-Factor Promax Solution

Empirical Factor	Items (primary loading ≥ .29)	n items	Dominant a priori stage(s)
Factor 1 — Focused Reception	Q2–Q20 (19 items)	19	Receiving/Attending + Understanding
Factor 2 — Critical Evaluation	Q31–Q40, Q49	11	Evaluating
Factor 3 — Retention/Memory	Q21, Q23–Q25, Q27–Q30	8	Remembering
Factor 4 — Responsive Engagement I	Q1, Q22, Q26, Q41, Q44–Q47, Q50	9	Responding (+2 Remembering, +1 Receiving)
Factor 5 — Responsive Engagement II	Q42, Q43, Q48	3	Responding

The pattern in Table 2 shows partial, rather than complete, correspondence with the a priori five-stage design. The receiving/attending and understanding stages, hypothesized as separate dimensions, loaded onto a single broad Focused Reception factor, suggesting that Region II HEI respondents did not distinguish, in their self-reports, between attending to a speaker and comprehending the speaker's message — a plausible outcome given that both stages occur in rapid succession during real-time listening and may be difficult for respondents to introspectively separate. In contrast, the evaluating and remembering stages emerged as clean, largely unidimensional factors closely matching their hypothesized item sets. The responding stage split across two related factors (Factor 4 and Factor 5), both dominated by responding items but differing in emphasis — Factor

4 loading items describing embodied/nonverbal engagement and initiating verbal responses, and Factor 5 loading items describing follow-up questioning and constructive expression of disagreement. Six items (Q2, Q12, Q13, Q18, Q29, Q30) showed meaningful cross-loadings (> .32) on more than one factor and are flagged as candidates for revision or removal in future refinements of the scale.

4.3 Inter-Item and Inter-Scale Correlations

Table 3 summarizes mean inter-item correlations within each subscale, computed on the full sample (N = 314). Mean within-subscale inter-item correlations ranged from .476 (Receiving/Attending) to .598 (Evaluating), with an overall mean inter-item correlation of .453 across all 50 items. These values sit above the .15–.50 range conventionally recommended for unidimensional item sets (Clark & Watson, 1995) — a pattern consistent with the very high alpha coefficients reported in Section 4.6 and with the dominant general factor identified in the parallel analysis, and reinforcing the interpretation that several items within each subscale are highly redundant rather than each capturing a distinct facet of that stage.

Table 3. Mean Inter-Item Correlations Within Subscales

Subscale	k (items)	Mean r	Min r	Max r
Receiving/Attending	10	.476	.355	.572
Understanding	10	.522	.302	.689
Remembering	10	.515	.193	.644
Evaluating	10	.598	.487	.715
Responding	10	.547	.326	.692
Overall (50 items)	50	.453	.160	.715

Table 4 reports the inter-scale (composite-score) correlation matrix. All five subscales correlated strongly and positively with one another (r = .68–.82), with Evaluating and Responding showing the strongest association (r = .820) and Receiving/Attending and Evaluating the weakest (r = .677) — still a large effect by conventional standards. These uniformly strong positive correlations are consistent with a single, strong underlying “general listening” dimension running through all five hypothesized stages, foreshadowing the confirmatory and structural results reported next.

Table 4. Inter-Scale (Composite Score) Correlation Matrix

Subscale	Receiving/Attending	Understanding	Remembering	Evaluating	Responding
Receiving/Attending	—	.810	.728	.677	.730
Understanding	.810	—	.751	.750	.736
Remembering	.728	.751	—	.742	.747
Evaluating	.677	.750	.742	—	.820
Responding	.730	.736	.747	.820	—

4.4 Confirmatory Factor Analysis

To formally test the measurement structure suggested by the EFA, four nested models were estimated by maximum likelihood on the screened sample (N = 291) using semopy: (1) the a priori five-factor correlated model corresponding to the original subscale design; (2) an EFA-informed five-factor correlated model using the empirical item groupings in Table 2 (cross-loading items assigned to their highest-loading factor); (3) a second-order (hierarchical) model in which the five a priori first-order factors load onto a single higher-order General Listening Process factor; and (4) a full structural model extending Model 3 with a structural path from the General factor to a numeric approximation of academic performance (bracket midpoints: below 80 = 75, 80–85 = 82.5, 85–89 = 87, 90–100 = 95). Fit indices for all four models are reported in Table 5.

Table 5. Fit Indices for Confirmatory and Structural Models

Model	df	χ^2	p	CFI	TLI	RMSEA	AIC
Model 1: A priori 5-factor (correlated)	1165	2240.08	<.001	.883	.877	.056	204.60
Model 2: EFA-informed 5-factor (correlated)	1165	2326.55	<.001	.873	.867	.059	204.01
Model 3: Second-order (hierarchical) factor	1170	2305.35	<.001	.876	.870	.058	194.16
Model 4: Full structural model (+ GWA outcome)	1219	2371.69	<.001	.874	.869	.057	197.70

All four models produced RMSEA values in the "acceptable" range (.056–.059, below the .06–.08 conventional ceiling) but CFI and TLI values below the conventional .90–.95 thresholds (.873–.883), indicating adequate but not excellent absolute fit — an outcome broadly consistent with the merged and split factors already identified in the EFA. The a priori five-factor model (Model 1) showed marginally better fit by chi-square, CFI, and TLI than the EFA-informed model (Model 2), while the second-order model (Model 3) fit nearly identically to Model 1 despite being more parsimonious (5 additional degrees of freedom for essentially the same chi-square) and achieved the lowest AIC of the three measurement models. This pattern — a hierarchical model matching the fit of a freely correlated first-order model at lower complexity — is itself evidence for the strong general listening factor: standardized loadings of the five first-order factors onto the second-order General factor were uniformly high, ranging from .880 (Evaluating) to .912 (Understanding), indicating that all five hypothesized stages are strongly saturated by a common underlying listening competence. Standardized item loadings on their respective first-order factors in the a priori model ranged from .54 to .81 ($M = .71$), all well above the conventional .40 minimum for retention (Watkins, 2018).

4.5 Structural Equation Modeling: Listening Process and Academic Performance

Model 4 extended the second-order measurement model with a structural regression path from the General Listening Process factor to students' academic performance (operationalized as the numeric midpoint of their reported general-average bracket). This path was positive and statistically significant (unstandardized $B = 2.27$, standardized $\beta = .212$, $z = 3.38$, $p < .001$), indicating that a one-standard-deviation increase in the latent General listening factor was associated with roughly a 0.21-standard-deviation increase in academic performance, holding the measurement model constant. Overall model fit for the full structural model (Table 5, Model 4) was essentially unchanged from the pure measurement model (Model 3), confirming that the structural addition did not degrade fit. This structural result corroborates, within a full latent-variable framework, the nonparametric known-groups finding reported in Section 4.8 (Table 8) and provides the strongest single piece of evidence in this study that the LPS's General factor captures a construct with genuine academic relevance rather than a measurement artifact.

4.6 Internal Consistency Reliability

Cronbach's alpha for the full 50-item scale was .976 ($N = 314$), and alpha values for the five a priori subscales ranged from .900 to .937 (Table 6), all exceeding the conventional .80 threshold for research instruments (Taber, 2018). Corrected item-total correlations exceeded .30 for every item, and no subscale's alpha would have improved materially through removal of any single item, indicating that all 50 items contribute meaningfully to their respective subscales despite the more nuanced structure revealed by the factor analysis.

Table 6. Internal Consistency Reliability of the LPS

Subscale	k (items)	Cronbach's α	Interpretation
Receiving/Attending	10	0.900	Excellent
Understanding	10	0.914	Excellent
Remembering	10	0.913	Excellent
Evaluating	10	0.937	Excellent
Responding	10	0.922	Excellent
Overall Scale	50	0.976	Excellent

4.7 Descriptive Level of the Listening Process

Using the conventional five-point verbal-interpretation range (1.00–1.79 Very Low, 1.80–2.59 Low, 2.60–3.39 Moderate, 3.40–4.19 High, 4.20–5.00 Very High), respondents rated themselves at a "High" level on all five listening dimensions and overall (Table 7), with means ranging narrowly from 3.47 (Remembering) to 3.67 (Responding). The Remembering dimension had both the lowest mean and among the more restricted subscale reliabilities, consistent with retention being commonly reported as the most effortful stage of listening in prior literature (Wolvin & Coakley, 1996).

Table 7. Descriptive Levels of the Listening Process

Dimension	Mean	SD	Verbal Interpretation
Receiving/Attending	3.66	0.67	High
Understanding	3.55	0.64	High
Remembering	3.47	0.63	High
Evaluating	3.60	0.66	High
Responding	3.67	0.67	High
Overall Listening Process	3.59	0.58	High

4.8 Differences Across Profile Variables

Because Shapiro–Wilk testing indicated the overall composite score departed significantly from normality ($W = .977$, $p < .001$), nonparametric tests were used for group comparisons (Table 8). Female respondents reported significantly higher scores than male respondents on Understanding (Mdn difference, $p = .049$), Remembering ($p = .045$), Responding ($p = .017$), and the Overall composite ($p = .029$), though not on Receiving/Attending or Evaluating. Neither age bracket nor first-language background produced statistically significant differences on any dimension. Academic performance bracket, by contrast, was significantly associated with every listening dimension (all $p \leq .007$), with mean Overall LPS scores rising monotonically from 3.28 among students in the below-80 bracket to 3.51 (80–85), 3.51 (85–89), and 3.79 (90–100).

Table 8. Group Comparisons Across Profile Variables

Dimension	Sex (U)	Sex (p)	Age (H)	Age (p)	Language	GWA	GWA (p)	p-value
Receiving/Attending	10339.50	.116	13.19 $^{(ns)}$	.528	not sig.	not sig.	12.27	.007*
Understanding	10031.50	.049*	15.20 $^{(ns)}$	.604	not sig.	not sig.	14.60	.002*
Remembering	10009.00	.045*	16.63 $^{(ns)}$	.408	not sig.	not sig.	21.00	<.001*
Evaluating	10693.50	.261	14.19 $^{(ns)}$	.908	not sig.	not sig.	14.39	.002*
Responding	9711.00	.017*	13.16 $^{(ns)}$	.923	not sig.	not sig.	12.07	.007*
Overall	9859.00	.029*	13.74 $^{(ns)}$	.691	not sig.	not sig.	17.09	.001*

This monotonic association between self-reported listening scores and academic performance provides meaningful criterion-related validity evidence: a scale capturing a construct genuinely relevant to classroom learning would be expected to covary with an independent academic outcome, and it does so consistently across every hypothesized dimension of the instrument.

5. Discussion

This study set out to validate a stage-based measure of the listening process for HEI students in Region II, Philippines, and to describe the level at which students engage in each stage. Four findings merit discussion.

First, the LPS demonstrated excellent internal consistency at both the subscale and full-scale level, and excellent sampling adequacy for factor analysis — indicating that the item pool is coherent and well-suited to measuring self-reported listening behavior in this population. The very high overall alpha (.976), however, should be interpreted cautiously rather than celebrated uncritically: alpha values in this range can reflect item redundancy as much as true reliability (Taber, 2018), and the strong general factor identified in the parallel analysis is consistent with this interpretation. Future revisions of the LPS might consider trimming redundant items within each subscale without materially reducing reliability, thereby producing a shorter, more efficient instrument.

Second, the empirical factor structure diverged from the a priori five-stage design in an interpretable way. The receiving/attending and understanding stages merged into a single Focused Reception factor, while evaluating and remembering emerged as clean, distinct factors, and responding split into two related facets. This pattern echoes a recurring finding in the broader listening-measurement literature, where self-report listening instruments tend to produce a dominant general factor alongside a smaller number of more specific facets rather than cleanly separable stage-by-stage factors (Bodie, Worthington, Imhof, & Cooper, 2008; Bodie, 2013). One plausible explanation is that receiving and understanding occur nearly simultaneously during real-time listening, making them difficult for respondents to introspectively disentangle when rating their own behavior, whereas remembering (a later, more effortful and consciously monitored process) and evaluating (a more deliberate, judgment-based process) are more psychologically distinguishable and therefore easier to rate as separate constructs. This suggests that instrument developers targeting the early stages of the listening process may need either more behaviorally granular items or a different measurement approach (e.g., performance-based rather than self-report tasks) to achieve empirical separation.

Third, the confirmatory and structural analyses converge with, and extend, the exploratory findings. None of the four CFA/SEM models reached conventional "good fit" thresholds on CFI or TLI, even though RMSEA fell in the acceptable range for all of them — a common pattern for long, highly intercorrelated Likert scales, and one that should temper any claim that the LPS's factor structure is fully confirmed. More informative than any single model's fit, however, is the comparison across models: a second-order model in which all five stages loaded onto one General Listening Process factor matched the fit of the more complex, freely correlated first-order model at a lower AIC, and every first-order factor loaded on that general factor at .88 or above. Read together with the uniformly high inter-scale correlations (Table 4) and the elevated inter-item correlations within every subscale (Table 3), this is convergent, multi-method evidence — from EFA, CFA, and inter-item correlation analysis alike — that respondents' self-reported listening behavior is best described as one dominant listening-competence dimension expressed through five correlated facets, rather than five fully independent abilities. This has a direct practical implication: institutions adopting the LPS should treat the Overall/General score as the primary, best-supported unit of measurement, using the five subscale scores as diagnostic supplements rather than as independently validated stand-alone measures until further refinement reduces item redundancy and cross-loading.

Fourth, the known-groups and structural analyses offer encouraging, if partial, validity evidence. The monotonic relationship between academic performance bracket and every LPS dimension — corroborated by the significant structural path from the latent General factor to academic performance in the full SEM ($\beta = .212$, $p < .001$) — is the strongest evidence in this study that the scale captures something practically consequential rather than an artifact of item wording. The sex difference — female respondents scoring higher on three of five dimensions and overall — is consistent with prior communication research reporting modestly higher self-reported listening or interpersonal-communication competence among female respondents (Bodie, 2011), though self-report differences of this kind may also reflect response-style differences between groups rather than differences in underlying listening skill, a distinction this design cannot resolve. The absence of differences by age bracket or first-language background suggests that, within this sample, the listening process as measured by the LPS operates comparably across these strata.

Practically, these findings suggest that HEI faculty and student-support units in Region II can use the LPS as a diagnostic screening tool, particularly for identifying students who may benefit from targeted support in retention (Remembering) or critical evaluation, the two stages that emerged as the cleanest, most distinguishable empirical dimensions and therefore the most defensible for individual-level score interpretation. Interpreting Receiving/Attending and Understanding as a single "Focused Reception" composite, rather than as two separate scores, is recommended until confirmatory work establishes otherwise.

6. Limitations

Several limitations qualify these findings. First, the sample was recruited through convenience, online sampling rather than probability sampling, which limits generalizability to the broader population of Region II HEI students and may have contributed to the data-quality concerns (23 straight-lining respondents) documented in Section 3.4. Second, all data are self-reported; the LPS measures students' perceptions of their own listening behavior, which may diverge from observed or performance-based listening competence. Third, although this study went beyond exploratory factor analysis to test the measurement structure with confirmatory factor analysis and a structural equation model, all four CFA/SEM models were estimated on the same sample used to inform the EFA-based model (Model 2) and, for Model 1, the same sample used to design the a priori subscales; independent-sample cross-validation, ideally with a second Region II or national cohort, is needed before treating the confirmatory fit indices as fully out-of-sample evidence. Relatedly, the ratio of sample size (N = 291) to freely estimated parameters in the 50-item CFA/SEM models (roughly 2.5:1) falls well short of common recommendations of 10:1 or higher, which can make chi-square and standard errors less trustworthy even though CFI, TLI, and RMSEA are comparatively robust to this concern; the fit indices reported here should accordingly be treated as provisional pending replication on a substantially larger sample. Fourth, test-retest reliability and convergent validity against an established listening instrument (e.g., the Watson–Barker Listening Test) were not assessed in this study and represent a priority for follow-up validation work. Fifth, the numeric academic-performance variable used in the structural model was approximated from categorical general-average brackets using bracket midpoints, a coarse proxy for actual grade point average that likely attenuates the true association between listening and academic performance. Finally, the cross-sectional, correlational design of the known-groups and structural analyses cannot establish that stronger listening skills cause better academic performance; the association is consistent with, but does not confirm, a causal relationship in either direction.

7. Conclusion and Recommendations

The Listening Process Scale demonstrated excellent internal consistency, strong sampling adequacy, and a coherent, hierarchically-organized factor structure — substantiated across exploratory factor analysis, inter-item/inter-scale correlation analysis, confirmatory factor analysis, and structural equation modeling — when validated on a sample of 314 (291 after data-quality screening) HEI students in Region II, Philippines. Students reported engaging in all five stages of the listening process at a consistently “High” level, and a latent General Listening Process factor showed a significant structural path to academic performance, supporting the instrument’s practical relevance beyond simple mean-level description. The instrument’s Evaluating and Remembering subscales performed as the most empirically distinct within the EFA and are recommended as diagnostic scores for identifying stage-specific weaknesses; the Receiving/Attending and Understanding items are recommended for interpretation as a single composite pending further refinement; and the Responding items may benefit from a formal two-facet revision distinguishing embodied engagement cues from verbal follow-up behavior. Because none of the CFA/SEM models reached conventional good-fit thresholds on CFI/TLI, and because the strong second-order general factor indicates considerable shared variance across all five stages, the single Overall/General score is currently the most defensible unit for high-stakes or comparative use, with the five subscale scores serving a formative, diagnostic role.

Future research should pursue confirmatory factor analysis and structural equation modeling on an independent Region II or nationally representative sample — ideally with a larger N to raise the parameter-to-sample ratio beyond that available here — establish convergent validity against an existing listening measure, assess test-retest reliability, and explore measurement invariance across sex and institution type. Item reduction targeting the six flagged cross-loading items (Q2, Q12, Q13, Q18, Q29, Q30) is a natural next step toward a shorter instrument with cleaner simple structure. Researchers and institutions adopting the LPS are encouraged to retain the data-quality screening procedures documented here, given the meaningful proportion of straight-lining responses observed in this online-administered sample.

References

Aguacito, R. G., Macatabon, R. A., & Imbang-Agni, E. J. D. (2025). Factors affecting the listening skills of college students. *E-DAWA, 5*(2), 1–15. University of Southern Mindanao.

Bodie, G. D. (2011). The Active-Empathic Listening Scale (AELS): Conceptualization and evidence of validity within the interpersonal domain. *Communication Quarterly, 59*(3), 277–295. https://doi.org/10.1080/01463373.2011.583495

Bodie, G. D. (2013). Issues in the measurement of listening. *Communication Research Reports, 30*(1), 76–84. https://doi.org/10.1080/08824096.2012.733981

Bodie, G. D., & Worthington, D. L. (2010). Revisiting the Listening Styles Profile (LSP-16): A confirmatory factor analytic approach to scale validation and reliability estimation. *International Journal of Listening, 24*(2), 69–88. https://doi.org/10.1080/10904011003744516

Bodie, G. D., Worthington, D. L., & Fitch-Hauser, M. (2011). A comparison of four measurement models for the Watson–Barker Listening Test (WBLT)-Form C. *Communication Research Reports, 28*(1), 32–42. https://doi.org/10.1080/08824096.2011.540547

Bodie, G. D., Worthington, D. L., & Gearhart, C. C. (2013). The Listening Styles Profile-Revised (LSP-R): A scale revision and evidence for validity. *Communication Quarterly, 61*(1), 75–93. https://doi.org/10.1080/01463373.2012.720343

Bodie, G. D., Worthington, D. L., Imhof, M., & Cooper, L. O. (2008). What would a unified field of listening look like? A proposal linking past perspectives and future endeavors. *International Journal of Listening,* *22*(2), 103–122. https://doi.org/10.1080/10904010802174867

Brownell, J. (1985). A model for listening instruction: Management applications. *Bulletin of the Association for Business Communication, 48*(3), 39–44.

Brownell, J. (1996). *Listening: Attitudes, principles, and skills*. Allyn & Bacon.

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment, 7*(3), 309–319. https://doi.org/10.1037/1040-3590.7.3.309

Gilakjani, A. P., & Ahmadi, M. R. (2011). A study of factors affecting EFL learners' English listening comprehension and the strategies for improvement. *Journal of Language Teaching and Research, 2*(5), 977–988.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Igolkina, A. A., & Meshcheryakov, G. (2020). semopy: A Python package for structural equation modeling. *Structural Equation Modeling, 27*(6), 952–963. https://doi.org/10.1080/10705511.2019.1704289

Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods, 17*(3), 437–455. https://doi.org/10.1037/a0028085

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

Watkins, M. W. (2018). Exploratory factor analysis: A guide to best practice. *Journal of Black Psychology, 44*(3), 219–246. https://doi.org/10.1177/0095798418771807

Wolvin, A. D., & Coakley, C. G. (1996). *Listening* (5th ed.). Brown & Benchmark.