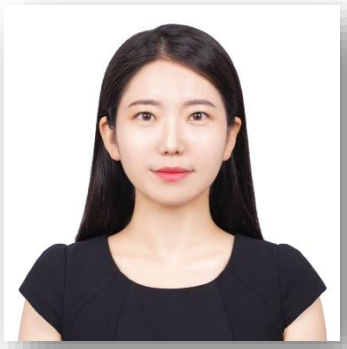
The International Research Foundation
for English Language Education

Researcher
Jieun Kim
University of Hawai'i Mānoa



Title of Project
Handwritten Versus Typed Notes: The Impact of Note-Taking Modes in Second Language Listening Tests

Research Supervisor
Prof. Daniel Isbell
University of Hawai'I Mānoa

TIRF Research Topic Investigated
Language Assessment

Bio
Jieun Kim holds a PhD from the University of Hawai'i at Mānoa, focusing on language testing. She taught English in Korea and received her MA from Seoul National University. Her dissertation examines the effects of different note-taking modes (handwriting versus typing) on L2 learners' listening comprehension and contents of notes (e.g., word count, translanguaging).

Final Report

Motivation for the Research
Assessing academic listening is crucial in language tests for university admissions, as it reflects real-world academic demands. Students often take notes during lectures, a practice shown to aid learning in both first and second language (L1 and L2) contexts. However, the effectiveness of note-taking varies based on its purpose, the content of the notes (e.g., word count, organization, use of symbols, verbatim vs. paraphrased text), and whether students switch languages in their notes.

With more students using computers in classrooms, typing is becoming more common. Typing generally results in longer, more verbatim notes but may lack nonlinguistic elements and deeper processing compared to handwriting. These shifts have implications for language assessments, many of which are now computerized. Note-taking policies in these tests vary widely and are often shaped by logistical concerns rather than research evidence.

Currently, there is limited research on how different note-taking methods affect listening test performance. Only one L2 study (Cubilo, 2017) has explored this topic, and it lacked a no note-taking condition and detailed content analysis of the notes. Addressing this gap, the present dissertation investigates how different note-taking modes—handwriting, typing, and no note-taking—affect L2 learners' performance on lecture comprehension tasks in standardized English proficiency tests. It also analyzes note quality across modes, aiming to inform more equitable and evidence-based testing policies.

Research Questions

1. To what extent do learners' listening test performances differ based on note-taking modes (i.e., handwriting, typing, and no note-taking)?
2. To what extent do test takers' note quality differ between handwritten notes and typed notes?
3. Which note quality measures accurately predict listening comprehension?

Research Methodology

Participants

The participants in the present study comprised 305 EFL Korean university students, including 205 females and 99 males, with one participant choosing not to disclose their gender. Their age ranged from 18 to 52 years, with a mean age of 23.84 years (*SD* = 4.87, median = 23). The oldest participant, who was a 52-year-old graduate student, was planning to study in the U.S. in the near future. Among the final participants, 221 were undergraduates, 18 were graduate students, three planned to enter university in the following academic year, and 29 had graduated from college and were either seeking employment or planning to apply for graduate programs. The remaining 34 were professionals who had graduated from 4-year universities.

Among the final 305 participants, 101 participants were assigned to the control condition, 102 to the handwriting condition, and 102 to the typing condition. Based on self-assessed English proficiency, the three note-taking groups exhibited a similar distribution.

Instruments

Four sets of lectures from the TOEFL iBT, each accompanied by six multiple-choice items, were utilized in this study. To reduce the possibility that Korean participants were familiar with freely accessible practice items and commercial materials, the study used items that were not publicly available. These items were obtained from the TOEFL developer, Educational Testing Service (ETS), through a data request. The mean length of a listening passage was approximately five minutes, with an average of 774 words and a speech rate of 149 words per minute. The comprehension questions were presented after each listening passage. Each set included six multiple-choice items targeting the main idea, supporting details, or inferences. Most items had a single correct answer, but four required selecting two or three correct options. The reliability was Cronbach's alpha = .83 and McDonald's omega = .84. Although the present study included only a single item type, this reliability is consistent with the overall listening section reliability of .83 reported by ETS (2020).

Procedure

The experiment was conducted in person in a quiet room, with a maximum of six participants at a time. Participants were randomly assigned to one of three groups: no note-taking, handwriting, or typing. Participants tested at the same time were assigned to the same note-taking condition. They began by providing consent and accessing the experiment link provided by the researcher. Upon accessing the link, participants completed four TOEFL lecture comprehension tasks, which involved listening to an audio recording while viewing a related image. The audio played automatically once, followed by six multiple-choice questions related to the content. Participants were not allowed to preview the test items, consistent with TOEFL guidelines and Cubilo's (2017) procedures.

Analysis

To address Research Question 1, the overall effects of note-taking mode were analyzed using regression analysis. The independent variable was note-taking group (with three levels), and the dependent variable was the total score out of 24 (4 lectures × 6 items). The alpha level was set at .05. In addition, item-level effects were examined using Rasch modeling with a Differential Item Functioning (DIF) approach. DIF analyses were conducted using Winsteps version 5.5.1, which supports both uniform and non-uniform DIF detection.

To address Research Question 2, note quality was assessed based on several dimensions: word count, information level, verbatim writing, translanguaging, and nonlinguistic elements. Word count was simply tallied for each note. Information units were identified and compared to those in the listening passages in terms of number and level (e.g., major vs. minor ideas). Verbatim writing was assessed by calculating n-gram overlap and stem overlap between the notes and the passages, following prior studies (Luo et al., 2018; Morehead et al., 2019; Mueller & Oppenheimer, 2014). For translanguaging, the percentage of non-English words in the notes was calculated. Nonlinguistic elements such as arrows and drawings were counted manually, particularly for handwritten notes. As normality assumptions were not met for all the note quality measures, the nonparametric Mann-Whitney U test was used. All statistical analyses were conducted in R version 4.3.1 (R Core Team, 2023).

Summary of Findings

Test scores were numerically the highest in the handwriting condition, followed by typing and no note-taking. However, this difference was not significant based on the linear regression. Uniform DIF analyses revealed that one item requiring inference favored handwriting, while another item focusing on main ideas favored typing. No DIF was observed when comparing no note-taking with handwriting or typing. Non-uniform DIF emerged only among lower-ability test takers, with the same items flagged in uniform DIF analyses showing discrepancies between handwriting and typing.

Analysis of note content showed no significant differences between handwritten and typed notes in terms of the number of words (W = 4,802, p = .474, r = -0.06 [95% CI -0.21, 0.10]) and translanguaging (the proportion of non-English words) (W = 4,649.5, p = .278, r = .09, [95% CI -0.24, 0.07]). However, handwritten notes included more information units from the passage (W = 5,923.5, p = .048, r = .16 [95% CI 0.00, 0.31]), verbatim text from the passage (W = 6,202, p = .024, r = .22, [95% CI 0.06, 0.36]), and nonlinguistic elements (W = 8,804.5, p < .001, r = .73, [95% CI 0.64, 0.79])than typed notes.

Based on regression analyses, for the handwriting group, information units were the only significant predictor of test scores while for the typing group, information units positively predicted scores. Verbatim writing and translanguaging had negative impacts.

In summary, the primary finding of this study is the absence of a significant difference in listening test performance across different note-taking modes—control, handwriting, and typing. However, this result should be interpreted in light of the participant population, study materials, test tasks, and the test-taking context. Although no significant differences were found in total listening test scores across note-taking modes, some differences did emerge at the item level.

Furthermore, notes taken in the handwriting and typing conditions differed in terms of information units, verbatim writing, and nonlinguistic elements. These differences raise questions about how they might influence not just test scores, but also deeper aspects of processes underlying listening comprehension that go beyond what is reflected in the overall test scores.

Implications

This study's findings—particularly the non-significant differences in listening comprehension across handwriting, typing, and no note-taking modes—carry important implications for standardized language test developers. While current note-taking policies often prioritize practical and security concerns, they should also be grounded in empirical evidence to support test validity and fairness.

From a validity standpoint, note-taking policies relate to how well a test represents real-world academic tasks (domain definition), whether scores reflect the intended constructs (explanation), and the broader impact of test use (consequences). For example, if listening in English-medium universities involves diverse note-taking behaviors—including typing—then test tasks should reflect this diversity. Moreover, listening processes may subtly vary by note-taking mode, raising questions about whether certain policies (e.g., typing only) introduce unintended construct-irrelevant variance. While excluding note-taking may enhance security and practicality, it may also cause negative washback by discouraging a widely used academic skill.

Fairness concerns also arise. Inconsistent note-taking policies across test formats (e.g., handwriting on paper vs. typing online) may compromise score comparability if the target construct includes note-taking-related processes. Additionally, restricting note-taking to one mode may disadvantage test takers with disabilities or limited access to technology. Typing, in particular, can introduce digital literacy or L2 typing skills as a confounding factor, especially for those unfamiliar with English keyboard input.

This study suggests five possible policy directions: (a) excluding note-taking entirely; (b) allowing only handwriting; (c) allowing only typing; (d) matching note-taking modes to test delivery format (e.g., handwriting in test centers, typing online); and (e) offering both modes regardless of format. Each has trade-offs regarding construct representation, fairness, practicality, washback, and security.

Given the growing diversity of test delivery formats and learners' needs, test developers are encouraged to revisit current note-taking policies using evidence-based approaches. Allowing flexible note-taking modes may better align with modern academic practices, minimize construct-irrelevant variance, and support more inclusive, valid, and fair assessments.

References

Abbott, M. L. (2007). A confirmatory approach to differential item functioning on an ESL reading assessment. *Language Testing, 24*(1), 7–36. https://doi.org/10.1177/0265532207071510

ACTFL. (n.d.). Assigning CEFR Ratings to ACTFL Assessments. Retrieved September 20, 2024 from https://www.actfl.org/uploads/files/general/Assigning_CEFR_Ratings_To_ACTFL_Assessments.pdf

Aiken, E. G., Thomas, G. S., & Shennum, W. A. (1975). Memory for a lecture: Effects pf Notes, lecture rate, and informational density. *Journal of Educational Psychology, 67*(3), 439-444. https://doi.org/10.1037/h0076613

Aguilar-Roca, N. M., Williams, A. E., & O'Dowd, D. K. (2012). The impact of laptop-free zones on student performance and attitudes in large lectures. *Computers & Education, 59*(4), 1300-1308. https://doi.org/10.1016/j.compedu.2012.05.002

Alderson, J. C. & Bachman, L. F. (2001). Series editor's preface. In G. Buck (Ed.), *Assessing listening*. Cambridge University Press.

Allen, M., LeFebvre, L., LeFebvre, L., & Bourhis, J. (2020). Is the pencil mightier than the keyboard? A meta-analysis comparing the method of notetaking outcomes. *Southern Communication Journal, 85*(3), 143-154. https://doi.org/10.1080/1041794X.2020.1764613

American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME]. (2014). *The standards for educational and psychological testing*. AERA, APA, & NCME. https://www.testingstandards.net/open-access-files.html

Amini Asl, Z., & Kheirzadeh, S. (2016). The effect of note-taking and working memory on Iranian EFL learners' listening comprehension performance. *International Journal of Research Studies in Psychology, 5*(4), 41-51.

Anderson, J. R. (1980). Concepts, propositions, and schemata: What are the cognitive units. *Proceedings of Nebraska Symposium on Motivation, 28*, 121-162. https://apps.dtic.mil/sti/citations/tr/ADA085997

Anderson, R. C., & Carter, J. F. (1972). Retroactive inhibition of meaningfully-learned sentences. *American Educational Research Journal, 9*(3), 443-448. https://doi.org/10.3102/00028312009003443

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2019). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods, 52*(1), 388–407. https://doi.org/10.3758/s13428-019-01237-x

Aragón-Mendizábal, E., Delgado-Casas, C., Navarro-Guzmán, J. I., Menacho-Jiménez, I., & Romero-Oliva, M. F. (2016). A comparative study of handwriting and computer typing in note-taking by university students. *Communicar, 48*(24).

Aryadoust, V. (2020). Dynamics of item reading and answer changing in two hearings in a computerized while-listening performance test: An eye-tracking study. *Computer Assisted Language Learning, 33*(5-6), 510-537. https://doi.org/10.1080/09588221.2019.1574267

Aryadoust, V., Foo, S., & Ng, L. Y. (2022). What can gaze behaviors, neuroimaging data, and test scores tell us about test method effects and cognitive load in listening assessments?. *Language Testing, 39*(1), 56-89. https://doi.org/10.1177/0265532221102687

Bachman, L. F., & Palmer, A. S. (2010). *Language assessment in practice: Developing language assessments and justifying their use in the real world*. Oxford University Press.

Bae, J., & Bachman, L. F. (1998). A latent variable approach to listening and reading: Testing factorial invariance across two groups of children in the Korean/English two-way immersion program. *Language Testing, 15*(3), 380–414. https://doi.org/10.1177/026553229801500304

Badger, R., White, G., Sutherland, P., & Haggis, T. (2001). Note perfect: an investigation of how students view taking notes in lectures. *System, 29*(3), 405-417. https://doi.org/10.1016/S0346-251X(01)00028-8

Baker, L., & Lombardi, B. R. (1985). Students' lecture notes and their relation to test performance. *Teaching of Psychology, 12*(1), 28-32. https://doi.org/10.1207/s15328023top1201_9

Bandalos, D. L. (2018). *Measurement theory and applications for the social sciences*. The Guilford Press.

Banerjee, J., & Papageorgiou, S. (2016). What's in a topic? Exploring the interaction between test-taker age and item content in high-stakes testing. *International Journal of Listening, 30*(1–2), 8–24. https://doi.org/10.1080/10904018.2015.1056876

Barbier, M. L., Roussey, J. Y., Piolat, A., & Olive, T. (2006). Note-taking in second language: Language procedures and self-evaluation of the difficulties. *Current Psychology Letters. Behaviour, Brain & Cognition, 20*(3). https://doi.org/10.4000/cpl.1283

Barnett, J. E., Di Vesta, F. J., & Rogozinski, J. T. (1981). What is learned in note taking?. *Journal of Educational Psychology, 73*(2), 181. https://doi.org/10.1037/0022-0663.73.2.181

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255-278. https://doi.org/10.1016/j.jml.2012.11.001

Bartoń, K. (2024). MuMIn: Multi-Model Inference (Version 1.48.4) [R package]. CRAN. https://doi.org/10.32614/CRAN.package.MuMIn

Batty, A. O. (2015). A comparison of video-and audio-mediated listening tests with many-facet Rasch modeling and differential distractor functioning. *Language Testing, 32*(1), 3-20. https://doi.org/10.1177/0265532214531254

Berrío, Á. I., Gomez-Benito, J., & Arias-Patiño, E. M. (2020). Developments and trends in research on methods of detecting differential item functioning. *Educational Research Review, 31*, 100340. https://doi.org/10.1016/j.edurev.2020.100340

Blasiman, R. N., Dunlosky, J., & Rawson, K. A. (2017). The what, how much, and when of study strategies: Comparing intended versus actual study behaviour. *Memory, 25*(6), 784–792. https://doi.org/10.1080/09658211.2016.1221974

Bozorgian, H., & Pillay, H. (2013). Enhancing foreign language learning through listening strategy delivered in L1: an experimental study. *International Journal of Instruction, 6*(1). https://dergipark.org.tr/en/pub/eiji/issue/5138/70019

Bretzing, B. H., & Kulhavy, R. W. (1979). Notetaking and depth of processing. *Contemporary Educational Psychology, 4*(2), 145-153. https://doi.org/10.1016/0361-476X(79)90069-9

Brodersen, J., Meads, D., Kreiner, S., Thorsen, H., Doward, L., & McKenna, S. (2007). Methodological aspects of differential item functioning in the Rasch model. *Journal of Medical Economics, 10*(3), 309-324. https://doi.org/10.3111/13696990701557048

Broersma, M., & Cutler, A. (2011). Competition dynamics of second-language listening. *Quarterly Journal of Experimental Psychology, 64*(1), 74-95. https://doi.org/10.1080/17470218.2010.499174

Brownell, J. (2009). *Listening: Attitudes, principles, and skills*. (4th ed.). Allyn & Bacon.

Brunfaut, T. (2016). Assessing listening. In D. Tsagari & J. Banerjee (Eds.), *Handbook of second language assessment* (pp. 97-112). De Gruyter Mouton.

Bui, D. C., Myerson, J., & Hale, S. (2013). Note-taking with computers: Exploring alternative strategies for improved recall. *Journal of Educational Psychology, 105*(2), 299. https://doi.org/10.1037/a0030367

Bui, D. C., & Myerson, J. (2014). The role of working memory abilities in lecture note-taking. *Learning and Individual Differences, 33*, 12-22. https://doi.org/10.1016/j.lindif.2014.05.002

Burstein, J., LaFlair, G. T., Kunnan, A. J., & von Davier, A. A. (2022). A theoretical assessment ecosystem for a digital-first assessment—The Duolingo English Test. *Duolingo Research Report*, *22*(01). http://duolingo-testcenter.s3.amazonaws.com/media/resources/A+Theoretical+Assessment+Ecosystem+for+a+Digital-First+Assessment—The+Duolingo+English+Test.pdf

CAEL (n.d.). *What to expect on test day*. CAEL. Retrieved October, 2024 from https://www.cael.ca/what-to-expect-on-test-day/

CAEL (n.d.). *CAEL Online requirements checklist*. CAEL. Retrieved October, 2024 from https://www.cael.ca/wp-content/uploads/2022/01/CAEL-Online-Checklist-January2022.pdf

Canagarajah, S. (2011). Codemeshing in academic writing: Identifying teachable strategies of translanguaging. *The Modern Language Journal, 95*(3), 401-417. https://doi.org/10.1111/j.1540-4781.2011.01207.x

Carrell, P. L., Dunkel, P., & Mollaun, P. (2002). *The effects of notetaking, lecture length, and topic on the listening component of TOEFL 2000* (Vol. 23). Educational Testing Service.

Carrell, P. L., Dunkel, P. A., & Mollaun, P. (2004). The effects of notetaking, lecture length, and topic on a computer-based test of ESL listening comprehension. *Applied Language Learning, 14*(1), 83-105.

Carrell, P. L. (2007). Notetaking strategies and their relationship to performance on listening comprehension and communicative assessment tasks. *ETS Research Report Series, 2007*(1), i-60. https://doi.org/10.1002/j.2333-8504.2007.tb02043.x

Carter, J. F., & Van Matre, N. H. (1975). Note taking versus note having. Journal of Educational *Psychology, 67*(6), 900. https://doi.org/10.1037/0022-0663.67.6.900

Cenoz, J., & Gorter, D. (2017). Translanguaging as a pedagogical tool in multilingual education. In J. Cenoz, D. Gorter, & S. May (Eds.), *Language awareness and multilingualism* (3rd ed., pp. 309-321). Springer.

Chapelle, C. A. (2020). *Validity in language assessment*. Routledge.

Charness, G., Gneezy, U., & Kuhn, M. A. (2012). Experimental methods: Between-subject and within-subject design. *Journal of Economic Behavior & Organization, 81*(1), 1-8. https://doi.org/10.1016/j.jebo.2011.08.009

Chaudron, C., Cook, J., & Loschky, L. (1988). *Quality of lecture notes and second language listening comprehension* (No. 7). Center for Second Language Classroom Research, Social Science Research Institute, University of Hawaii at Manoa.

Chaudron, C., Loschky, L., & Cook, J. (1994). Second language listening comprehension and lecture note-taking. In J. Flowerdew (Ed.), *Academic listening: Research perspectives* (75-92). Cambridge University Press. https://doi.org/10.1017/CBO9781139524612.008

Chen, X., Aryadoust, V., & Zhang, W. (2024). A systematic review of differential item functioning in second language assessment. *Language Testing*, 02655322241290188. https://doi.org/10.1177/02655322241290188

Cho, D. (2023). *Hackers TOEFL listening*. Hackers Ehak Yenkwuso.

Cho, H. C., & Abe, S. (2013). Is two-tailed testing for directional research hypotheses tests legitimate?. *Journal of Business Research*, *66*(9), 1261-1266. https://doi.org/10.1016/j.jbusres.2012.02.023

Cho, J. R., & Chen, H. C. (1999). Orthographic and phonological activation in the semantic processing of Korean Hanja and Hangul. *Language and Cognitive Processes*, *14*(5-6), 481-502. https://doi.org/10.1080/016909699386167

Choi, I. C. (2008). The impact of EFL testing on EFL education in Korea. *Language Testing*, *25*(1), 39–62. https://doi.org/10.1177/0265532207083744

Choi, Y. B. (2013). Path dependence and the Korean keyboard. *Journal of Economic Behavior & Organization*, *88*, 37-46. https://doi.org/10.1016/j.jebo.2011.05.009

Clark, M., Wayland, S., Osthus, P., Brown, K. G., Castle, S., & Ralph, A. (2014). *The effects of note taking on foreign language listening comprehension*. University of Maryland Center for Advanced Study of Language.

Clerehan, R. (1995). Taking it down: Notetaking practices of L1 and L2 students. *English for Specific Purposes, 14*(2), 137-155. https://doi.org/10.1016/0889-4906(95)00003-A

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2013). *Applied multiple regression/correlation analysis for the behavioral sciences*. Routledge.

Conoley, C. (2003). *Differential item functioning in the Peabody Picture Vocabulary Test* [Unpublished doctoral dissertation]. Texas A&M University.

Crawford, C. C. (1925). Some experimental studies of the results of college note-taking. *The Journal of Educational Research, 12*(5), 379-386. https://doi.org/10.1080/00220671.1925.10879612

Crawford, M. (2015). A study on note taking in EFL listening instruction. *Proceedings of the JALT 2014*, Japan, 416-424. https://jalt-publications.org/sites/default/files/pdf-article/jalt2014proc_044.pdf

Cubilo, J. (2017). *Video-mediated listening passages and typed note-taking: Examining their effects on examinee listening test performance and item characteristics* [Unpublished doctoral dissertation]. University of Hawai'i at Manoa.

Cubilo, J., & Winke, P. (2013). Redefining the L2 listening construct within an integrated writing task: Considering the impacts of visual-cue interpretation and note-taking. *Language Assessment Quarterly, 10(*4), 371-397. https://doi.org/10.1080/15434303.2013.824972

Cutler, A. (2012). *Native listening: Language experience and the recognition of spoken words* (1st ed.). The MIT Press. https://doi.org/10.7551/mitpress/9012.001.0001

Cushing, S. T. (1991). The relationship between academic status, language proficiency and notetaking. *Unpublished manuscript, University of California, Los Angeles*.

Davidson, B. (2004). *Comparability of test scores for non-aboriginal and aboriginal students*. [Unpublished doctoral dissertation]. University of British Columbia.

Davis, M., & Hult, R. E. (1997). Effects of writing summaries as a generative learning activity during note taking. *Teaching of Psychology, 24*(1), 47-50. https://doi.org/10.1207/s15328023top2401_10

DeWinstanley, P. A., & Bjork, E. L. (2004). Processing strategies and the generation effect: Implications for making a better reader. *Memory & Cognition, 32*(6), 945-955. https://doi.org/10.3758/BF03196872

Di Vesta, F. J., & Gray, G. S. (1973). Listening and note taking: II. Immediate and delayed recall as functions of variations in thematic continuity, note taking, and length of listening-review intervals. *Journal of Educational Psychology, 64*(3), 278-287. https://doi.org/10.1037/h0034589

Dunkel, P. (1988). The content of L1 and L2 students' lecture notes and its relation to test performance. *TESOL Quarterly, 22*(2), 259-281. https://doi.org/10.2307/3586936

Dunkel, P., Mishra, S., & Berliner, D. (1989). Effects of note taking, memory, and language proficiency on lecture learning for native and nonnative speakers of English. *TESOL Quarterly, 23*(3), 543-549. https://doi.org/10.2307/3586929

Duolingo English Test. (n.d.). *Can I take notes during the test?* Duolingo English Test. Retrieved June 8, 2024, from https://testcenter.zendesk.com/hc/en-us/articles/360011087771-Can-I-take-notes-during-the-test

Duolingo English Test. (2024). *Official guide for test takers*. Duolingo English Test.

Einstein, G. O., Morris, J., & Smith, S. (1985). Note-taking, individual differences, and memory for lecture information. *Journal of Educational Psychology, 77*(5), 522. https://doi.org/10.1037/0022-0663.77.5.522

Elosua, P. (2024) A three-step DIF analysis of a reading comprehension test across regional dialects to improve test score validity. *Language Assessment Quarterly, 21*(2), 141-158. https://doi.org/10.1080/15434303.2024.2307621

ETS. (2019). *Mapping the TOEIC Tests on the CEFR*. ETS. https://www.ets.org/pdfs/toeic/toeic-mapping-cefr-reference.pdf

ETS. (2020). *Reliability and comparability of TOEFL iBT Scores*. ETS. https://www.ets.org/pdfs/toefl/toefl-ibt-insight-s1v3.pdf

ETS. (2022). *TOEIC® listening and reading test examinee handbook*. ETS. https://www.ets.org/pdfs/toeic/toeic-listening-reading-test-examinee-handbook.pdf

ETS. (n.d.). *Comparing TOEFL iBT Scores*. ETS. Retrieved September 20, 2024 from https://www.ets.org/toefl/institutions/ibt/compare-scores.html#accordion-ff2afbe1ab-item-95a6512adf

Faraco, M., Barbier, M., & Piolat, A. (2002). A comparison between notetaking in L1 and L2 by undergraduate students. In S. Ransdell & M. Barbier (Eds.), *Studies in writing, volume 11: New directions for research in L2 writing* (pp. 145–167). Kluwer Academic Publishers.

Ferne, T., & Rupp, A. A. (2007). A synthesis of 15 years of research on DIF in language testing: Methodological advances, challenges, and recommendations. *Language Assessment Quarterly, 4*(2), 113-148. https://doi.org/10.1080/15434300701375923

Feyten, C. M. (1991). The power of listening ability: An overlooked dimension in language acquisition. *The Modern Language Journal, 75*(2), 173-180. https://doi.org/10.2307/328825

Field, J. (2004). An insight into listeners' problems: Too much bottom-up or too much top-down?. *System, 32*(3), 363-377. https://doi.org/10.1016/j.system.2004.05.002

Field, J. (2013). Cognitive validity. In A. Geranpayeh & L. B. Taylor (Eds.), *Examining listening: Research and practice in assessing second language listening* (pp. 77-151). Cambridge University Press.

Fiorella, L., & Mayer, R. E. (2017). Spontaneous spatial strategy use in learning from scientific text. *Contemporary Educational Psychology, 49*, 66-79. https://doi.org/10.1016/j.cedpsych.2017.01.002

Fisher, J. L., & Harris, M. B. (1973). Effect of note taking and review on recall. *Journal of Educational Psychology, 65*(3), 321. https://doi.org/10.1037/h0035640

Flanigan, A. E., Wheeler, J., Colliot, T., Lu, J., & Kiewra, K. A. (2024). Typed versus handwritten lecture notes and college student achievement: A meta-analysis. *Educational Psychology Review*, *36*(3), 78. https://doi.org/10.1007/s10648-024-09914-w

Flowerdew, J., & Miller, L. (2010). Listening in a second language. In A. D. Wolvin (Ed.), *Listening and human communication in the 21st century* (pp. 158-177). Blackwell Publishing.

Frangou, S. M., Ruokamo, H., Parviainen, T., & Wikgren, J. (2018). Can you put your finger on it? The effects of writing modality on Finnish students' recollection. *Writing Systems Research, 10*(2), 82-94. https://doi.org/10.1080/17586801.2018.1536015

Frase, L. T. (1970). Boundary conditions for mathemagenic behaviors. *Review of Educational Research, 40*(3), 337-347. https://doi.org/10.3102/00346543040003337

Freedle, R., & Kostin, I. (1997). Predicting black and white differential item functioning in verbal analogy performance. *Intelligence, 24*(3), 417–444. https://doi.org/10.1016/S0160-2896(97)90058-1

Furuya, A. (2021). How do listening comprehension processes differ by second language proficiency? Top-down and bottom-up perspectives. *International Journal of Listening, 35*(2), 123-133. https://doi.org/10.1080/10904018.2019.1694411

García, O. (2009). Education, multilingualism and translanguaging in the 21st century. In T. Skutnabb-Kangas, R. Phillipson, A. Mohanty, & M. Panda (Eds.), *Social justice through multilingual education* (pp. 140-158). Multilingual Matters. https://doi.org/10.21832/9781847691910

Geramipour, M., & Shahmirzadi, N. (2004). A gender-related differential item functioning study of an English test. *The Journal of Asia TEFL, 16*(2), 674–682. http://dx.doi.org/10.18823/asiatefl.2019. 16.2.15.674

Gierl, M. J. (2005). Using dimensionality-based DIF analyses to identify and interpret constructs that elicit group differences. *Educational Measurement: Issues and Practice, 24*(1), 3–14. https://doi.org/10.1111/j.1745-3992.2005.00002.x

Glass, G. V. (1965). A ranking variable analogue of biserial correlation: Implications for short-cut item analysis. *Journal of Educational Measurement, 2*(1), 91-95. https://www.jstor.org/stable/1433839

Glass, G. V. (1966). Note on rank biserial correlation. *Educational and Psychological Measurement, 26*(3), 623-631. https://doi.org/10.1177/001316446602600307

Golia, S. (2015). Assessing the impact of uniform and nonuniform differential item functioning items on Rasch measure: The polytomous case. *Computational Statistics, 30*, 441-461. https://doi.org/10.1007/s00180-014-0542-x

Goodman, B., & Tastanbek, S. (2021). Making the shift from a codeswitching to a translanguaging lens in English language teacher education. *TESOL Quarterly, 55*(1), 29-53. https://doi.org/10.1002/tesq.571

Geary, D. (2012). Evolutionary educational psychology. In K. Harris, S. Graham, & T. Urdan (Eds.), *APA educational psychology handbook*, Vol. 1 (pp. 597–621). American Psychological Association. https://doi.org/10.1007/978-3-319-29986-0_9

Gurung, R. A. R. (2005). How do students really study (and does it matter)? *Teaching of Psychology, 32*(4), 239–241.

Hagquist, C., & Andrich, D. (2017). Recent advances in analysis of differential item functioning in health research using the Rasch model. *Health and Quality of Life Outcomes, 15*, 1-8. https://doi.org/10.1186/s12955-017-0755-0

Hale, G. A., & Courtney, R. (1994). The effects of note-taking on listening comprehension in the Test of English as a Foreign Language. *Language Testing, 11*(1), 29-47. https://doi.org/10.1177/026553229401100104

Hales, A. H. (2023). One-tailed tests: Let's do this (responsibly). *Psychological Methods*, *29*(6), 1209-1218. https://doi.org/10.1037/met0000610

Harding, L. (2012). Accent, listening assessment and the potential for a shared-L1 advantage: A DIF perspective. *Language Testing, 29*(2), 163–180. https://doi.org/10.1177/0265532211421161

Harding, L., Pill, J., & Ryan, K. (2011). Assessor decision making while marking a note-taking listening test: The case of the OET. *Language Assessment Quarterly, 8*(2), 108-126. https://doi.org/10.1080/15434303.2011.556770

Hartley, J., & Davies, I. K. (1978). Note-taking: A critical review. *Programmed Learning and Educational Technology, 15*(3), 207-224. https://doi.org/10.1080/0033039780150305

Hayati, A. M., & Jalilifar, A. (2009). The impact of note-taking strategies on listening comprehension of EFL learners. *English Language Teaching, 2*(1), 101-111. https://eric.ed.gov/?id=EJ1082250

Heppt, B., Haag, N., Böhme, K., & Stanat, P. (2015). The role of academic-language features for reading comprehension of language-minority students and students from low-SES families. *Reading Research Quarterly, 50*(1), 61–82. https://doi.org/10.1002/rrq.83

Huang, F. L. (2022). *Practical multilevel modeling using R*. Sage Publication.

idp IELTS. (n.d.). *IELTS on paper test*. idp IELTS. Retrieved June 9, 2024, from https://ielts.idp.com/about/which-test-do-i-take/paper-based-ielts

idp IELTS. (n.d.). *IELTS on computer test*. idp IELTS. Retrieved June 9, 2024, from https://ielts.idp.com/about/which-test-do-i-take/computer-delivered-ielts

IELTS. (n.d.). *IELTS Online*. IELTS. Retrieved June 8, 2024, from https://ielts.org/take-a-test/test-types/ielts-academic-test/ielts-online

IELTS. (n.d.). *IELTS and the CEFR*. IELTS. Retrieved September 20, 2024 from https://ielts.org/organisations/ielts-for-organisations/compare-ielts/ielts-and-the-cefr

Igo, L. B., Bruning, R., & McCrudden, M. A. (2005). Encoding disruption associated with copy and paste note taking. In L. M. PytilkZillig, M. Bodvarsson, & R. Bruning (Eds.), *Technology-based education: Bringing researchers and practitioners together* (pp.107-119). Information Age Publishing.

Isbell, D., Kremmel, B., Kim, J. (2023). Remote proctoring in language testing: Implications for fairness and justice. *Language Assessment Quarterly, 20*(4–5), 469–487. https://doi.org/10.1080/15434303.2023.2288251

James, K. H., & Engelhardt, L. (2012). The effects of handwriting experience on functional brain development in pre-literate children. *Trends in Neuroscience and Education, 1*(1), 32-42. https://doi.org/10.1016/j.tine.2012.08.001

James, K. H., & Gauthier, I. (2006). Letter processing automatically recruits a sensory–motor brain network. *Neuropsychologia, 44*(14), 2937-2949. https://doi.org/10.1016/j.neuropsychologia.2006.06.026

Jarmon, A. L. (2007). Multitasking-Helpful or Harmful. *Student Law, 36*, 31. https://heinonline.org/HOL/LandingPage?handle=hein.journals/studlyr36&div=57&id=&page=

Jeon, J. (2010). Issues for English tests and assessments: A view from Korea. In Y. Moon & B. Spolsky (Eds.), *Language assessment in Asia: Local, regional or global* (pp. 53–77). AisaTEFL Book Series. https://koreatesol.org/sites/default/files/pdf/(2)%20Jeon-A%20View%20from%20Korea%20(2010%20AsiaTEFL%20Book%20Series).pdf

Jin, Z., & Webb, S. (2024). The effectiveness of note taking through exposure to L2 input: A meta-analysis. *Studies in Second Language Acquisition, 46*(2), 404-426. https://doi.org/10.1017/S0272263123000529

Jurafsky, D., & Martin, J. H. (2008). *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition*. Pearson. https://www.dcs.bbk.ac.uk/~Dell/teaching/nlp/book/slp/slp_ch01.pdf

Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement, 50*(1), 1-73. https://doi.org/10.1111/jedm.12000

Kaushanskaya, M., & Yoo, J. (2013). Phonological short-term and working memory in bilinguals' native and second language. *Applied Psycholinguistics, 34*(5), 1005-1037. https://doi.org/10.1017/S0142716412000100

Kavak, V. İ., & Kırkgöz, Y. (2022). Exploring university students' note-taking in literature courses: A translanguaging perspective. *Çukurova University Faculty of Education Journal, 51*(2), 1468-1486. https://doi.org/10.14812/cuefd.1027947

Kerby, D. S. (2014). The simple difference formula: An approach to teaching nonparametric correlation. *Comprehensive Psychology, 3*. https://doi.org/10.2466/11.IT.3.1

Kiefer, M., Schuler, S., Mayer, C., Trumpp, N. M., Hille, K., & Sachse, S. (2015). Handwriting or typewriting? The influence of pen-or keyboard-based writing training on reading and writing performance in preschool children. *Advances in Cognitive Psychology, 11*(4), https://doi.org/136. 10.5709/acp-0178-7

Kiewra, K. A. (1985). Students' note-taking behaviors and the efficacy of providing the instructor's notes for review. *Contemporary Educational Psychology, 10*(4), 378-386. https://doi.org/10.1016/0361-476X(85)90034-7

Kılıçkaya, F., & Çokal-karadaş, D. (2009). The effect of note-taking on university students' listening comprehension of lectures. *Kastamonu Education Journal, 17*(1), 47-56. https://dergipark.org.tr/en/pub/kefdergi/issue/49070/626124

Kim, M. (2001). Detecting DIF across the different language groups in a speaking test. *Language Testing, 18*(1), 89-114. https://doi.org/10.1177/026553220101800104

Kim, K., Turner, S. A., & Pérez-Quiñones, M. A. (2009). Requirements for electronic note taking systems: A field study of note taking in university classrooms. *Education and Information Technologies, 14*, 255-283. https://doi.org/10.1007/s10639-009-9086-z

Kintsch, W. (1988). The role of knowledge in discourse comprehension: a construction-integration model. *Psychological Review, 95*(2), 163-182. https://doi.org/10.1037/0033-295X.95.2.163

Kobayashi, K. (2005). What limits the encoding effect of note-taking? A meta-analytic examination. *Contemporary Educational Psychology, 30*(2), 242-262. https://doi.org/10.1016/j.cedpsych.2004.10.001

Kobayashi, K. (2006a). Combined effects of note-taking/-reviewing on learning and the enhancement through interventions: A meta-analytic review. *Educational Psychology, 26*(3), 459-477. https://doi.org/10.1080/01443410500342070

Kobayashi, K. (2006b). Conditional effects of interventions in note-taking procedures on learning: A meta-analysis. *Japanese Psychological Research, 48*(2), 109-114. https://doi.org/10.1111/j.1468-5884.2006.00311.x

Kodaira, Y. (2017). *An investigation of cognitive processes associated with notetaking and notes-review* [Unpublished Doctoral dissertation]. Columbia University.

Kroll, J. F., Dussias, P. E., Bogulski, C. A., & Kroff, J. R. V. (2012). Juggling two languages in one mind: What bilinguals tell us about language processing and its consequences for cognition. In B. H. Ross (Ed.) *Psychology of learning and motivation* (Vol. 56, pp. 229-262). Academic press. https://doi.org/10.1016/B978-0-12-394393-4.00007-8

Kunnan, A. (2014). Fairness and Justice in Language Assessment. In A. Kunnan (Ed.) *The companion to language assessment* (pp. 1098-1114). John Wiley & Sons.

Kutta Jr, T. J. (2017). *Note-taking and technology* [Unpublished Doctoral dissertation]. The Florida State University.

Lakens, D. (2022). Sample size justification. *Collabra: Psychology*, *8*(1), 33267. https://doi.org/10.1525/collabra.33267

Langenfeld, T., Burstein, J., & von Davier, A. A. (2022). Digital-First learning and assessment systems for the 21st century. *Front. Educ*. *7*, 857606. https://doi.org/10.3389/feduc.2022.857604

Ladas, H. S. (1980). Note taking on lectures: An information-processing approach. *Educational Psychologist, 15*(1), 44-53. https://doi.org/10.1080/00461528009529215

Lai, J. S., Teresi, J., & Gershon, R. (2005). Procedures for the analysis of differential item functioning (DIF) for small sample sizes. *Evaluation & the Health Professions, 28*(3), 283-294. https://doi.org/10.1177/0163278705278276

Leonard, K. R. (2019). Examining the relationship between decoding and comprehension in L2 listening. *System, 87*, 102150. https://doi.org/10.1016/j.system.2019.102150

Li, Z., Chen, M. Y., & Banerjee, J. (2020). Using corpus analyses to help address the DIF interpretation: Gender differences in standardized writing assessment. *Frontiers in Psychology, 11*, 1088. https://doi.org/10.3389/fpsyg.2020.01088

Li, H., Hunter, C. V., & Bialo, J. A. (2022). A revisit of Zumbo's third generation DIF: How are we doing in language testing?. *Language Assessment Quarterly, 19*(1), 27-53. https://doi.org/10.1080/15434303.2021.1963253

Lin, M. C. (2006). *The effects of note-taking, memory, and rate of presentation on EFL learners' listening comprehension*. [Unpublished doctoral dissertation]. La Sierra University.

Linacre, J. M. (1994). Sample size and item calibration stability. *Rasch Measurement Transactions, 7*(4), 328. https://www.rasch.org/rmt/rmt74m.htm

Lee, B. J. (2021). Comparing factual recall of tapped vs. handwritten text. *Acta Psychologica, 212*, 103221. https://doi.org/10.1016/j.actpsy.2020.103221

Longcamp, M., Zerbato-Poudou, M. T., & Velay, J. L. (2005). The influence of writing practice on letter recognition in preschool children: A comparison between handwriting and typing. *Acta Psychologica, 119*(1), 67-79. https://doi.org/10.1016/j.actpsy.2004.10.019

Longcamp, M., Boucard, C., Gilhodes, J. C., & Velay, J. L. (2006). Remembering the orientation of newly learned characters depends on the associated writing knowledge: A comparison between handwriting and typing. *Human Movement Science, 25*(4-5), 646-656. https://doi.org/10.1016/j.humov.2006.07.007

Luo, L., Kiewra, K. A., Flanigan, A. E., & Peteranetz, M. S. (2018). Laptop versus longhand note taking: Effects on lecture notes and achievement. *Instructional Science, 46*, 947-971. https://doi.org/10.1007/s11251-018-9458-0

Lynch, T. (2011). Academic listening in the 21st century: Reviewing a decade of research. *Journal of English for Academic Purposes, 10*(2), 79-88. https://doi.org/10.1016/j.jeap.2011.03.001

Mangen, A., Anda, L. G., Oxborough, G. H., & Brřnnick, K. (2015). Handwriting versus keyboard writing: effect on word recall. *Journal of Writing Research, 7*(2), 227-247. https://doi.org/10.17239/jowr-2015.07.02.1

Malone, M. E., & Montee, M. (2014). Stakeholders' beliefs about the TOEFL iBT® test as a measure of academic language ability. *ETS Research Report Series*, *2014*(2), 1–51. https://doi.org/10.1002/ets2.12039

Mann, H. B., & Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics, 18*(1), 50-60. https://www.jstor.org/stable/2236101

Matthews, J., & Cheng, J. (2015). Recognition of high frequency words from speech as a predictor of L2 listening comprehension. *System, 52*, 1-13. https://doi.org/10.1016/j.system.2015.04.015

Marx, A., Heppt, B., & Henschel, S. (2017). Listening comprehension of academic and everyday language in first language and second language students. *Applied Psycholinguistics, 38*(3), 571-600. https://doi.org/10.1017/S0142716416000333

McNamara, T. F. (1990). *Assessing the second language proficiency of health professionals*. [Unpublished doctoral dissertation]. The University of Melbourne.

Messick, S. (1996). Validity and washback in language testing. *Language Testing, 13*(3), 241-256. https://doi.org/10.1177/026553229601300302

Morehead, K., Dunlosky, J., Rawson, K. A., Blasiman, R., & Hollis, R. B. (2019a). Note-taking habits of 21st century college students: implications for student learning, memory, and achievement. *Memory, 27*(6), 807-819. https://doi.org/10.1080/09658211.2019.1569694

Morehead, K., Dunlosky, J., & Rawson, K. A. (2019b). How much mightier is the pen than the keyboard for note-taking? A replication and extension of Mueller and Oppenheimer (2014). *Educational Psychology Review*, *31*, 753-780. https://doi.org/10.1007/s10648-019-09468-2

Mueller, P. A., & Oppenheimer, D. M. (2014). The pen is mightier than the keyboard: Advantages of longhand over laptop note taking. *Psychological Science, 25*(6), 1159-1168. https://doi.org/10.1177/0956797614524581

Mulligan, N. W. (2004). Generation and memory for contextual detail. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*(4), 838. https://doi.org/10.1037/0278-7393.30.4.838

Nakagawa, S., Johnson, P. C., & Schielzeth, H. (2017). The coefficient of determination R 2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *Journal of the Royal Society Interface*, *14*(134), 20170213. https://doi.org/10.1098/rsif.2017.0213

National Institute of Korean Language (국립국어원 kwuklipkwukewen). (2024). Loan words and foreign Language words(외국어 외래어 정의 oykwuke oylaye cenguy). Retrieved Jan 4, 2025, from https://www.korean.go.kr/front/onlineQna/onlineQnaView.do?mn_id=216&qna_seq=302466

Nishizawa, H. (2024). Authenticity of academic lecture passages in high-stakes tests: A temporal fluency perspective. *Language Testing*, *41*(4). https://doi.org/10.1177/02655322241262453

OET. (n.d.). *Can I make highlights, underline or write notes in the OET Test?*. OET. Retrieved June 8, 2024, from https://support.oet.com/s/article/Can-I-make-highlights-or-write-notes-in-OET-on-computer-at-a-venue#:~:text=Yes%2C%20you%20can%20make%20notes,a%20valuable%20test%2Dtaking%20strategy.

Olive, T., & Barbier, M. L. (2017). Processing time and cognitive effort of longhand note taking when reading and summarizing a structured or linear text. *Written Communication, 34*(2), 224-246. https://doi.org/10.1177/0741088317699898

Pae, H. K. (2018). Writing systems, reading processes, and cross-linguistic influences: Reflections from the Chinese, Japanese and Korean languages (1st ed., Vol. 7). John Benjamins Publishing Company. https://doi.org/10.1075/bpa.7

Pae, T.-I. (2004a). DIF for examinees with different academic backgrounds. *Language Testing, 21*(1), 53–73. https://doi.org/10.1191/0265532204lt274oa

Pae, H. K., Greenberg, D., & Morris, R. D. (2012). Construct validity and measurement invariance of the peabody picture vocabulary test–III form A. *Language Assessment Quarterly, 9*(2), 152–171. https://doi.org/10.1080/15434303.2011.613504

Palkovitz, R. J., & Lore, R. K. (1980). Note taking and note review: Why students fail questions based on lecture material. *Teaching of Psychology, 7*(3), 159-161. https://doi.org/10.1207/s15328023top0703_8

Park, H. (2019). The Effect of L1 and L2 Notetaking in Academic Lecture Listening on Listening Comprehension and Analysis of Notes. *Culminating Projects in TESL. 1*. https://repository.stcloudstate.edu/tesl_etds/1

Patterson, R. W., & Patterson, R. M. (2017). Computers and productivity: Evidence from laptop use in the college classroom. *Economics of Education Review, 57*, 66-79. https://doi.org/10.1016/j.econedurev.2017.02.004

Pearson. (n.d.). PTE help center: PTE academic online FAQs. Pearson. Retrieved June 8, 2024, from https://www.pearsonpte.com/help-center/general-faqs/pte-academic-online

Pearson. (2023). *PTE Academic Online Handbook*. Pearson. https://www.pearsonpte.com/ctf-assets/yqwtwibiobs4/78URBdlI1OHdcynWIiZqxC/cb8b7a685bde29e1f449ff8e79bcb3c0/PTE_Academic_Online_Test_Taker_Handbook_August_2023_V5.pdf

Pennycook, A. (2017). Translanguaging and semiotic assemblages. *International Journal of Multilingualism, 14*(3), 269-282. https://doi.org/10.1080/14790718.2017.1315810

Peper, R. J., & Mayer, R. E. (1986). Generative effects of note-taking during science lectures. *Journal of Educational psychology, 78*(1), 34. https://doi.org/10.1037/0022-0663.78.1.34

Perfetti. C. A. (1985). *Reading ability*. Oxford University Press.

Peters, D. L. (1972). Effects of note taking and rate of presentation on short-term objective test performance. *Journal of Educational Psychology, 63*(3), 276. https://doi.org/10.1037/h0032647

Peverly, S. T., Ramaswamy, V., Brown, C., Sumowski, J., Alidoost, M., & Garner, J. (2007). What predicts skill in lecture note taking?. *Journal of Educational Psychology, 99*(1), 167. https://doi.org/10.1037/0022-0663.99.1.167

Peverly, S. T., Vekaria, P. C., Reddington, L. A., Sumowski, J. F., Johnson, K. R., & Ramsay, C. M. (2013). The relationship of handwriting speed, working memory, language comprehension and outlines to lecture note-taking and test-taking among college students. *Applied Cognitive Psychology, 27*(1), 115-126. https://doi.org/10.1002/acp.2881

Plonsky, L., & Oswald, F. L. (2014). How big is "big"? Interpreting effect sizes in L2 research. *Language Learning,* 64(4), 878-912. https://doi.org/10.1111/lang.12079

Ponce, H. R., & Mayer, R. E. (2014). Qualitatively different cognitive processing during online reading primed by different study activities. *Computers in Human Behavior*, *30*, 121-130. https://doi.org/10.1016/j.chb.2013.07.054

Puhan, G., Boughton, K., & Kim, S. (2007). Examining Differences in Examinee Performance in Paper and Pencil and Computerized Testing. *Journal of Technology, Learning, and Assessment*, *6*(3), n3.

R Core Team (2023). R: A Language and Environment for Statistical Computing. Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org.

Rana, N. P., Slade, E., Kitching, S., & Dwivedi, Y. K. (2019). The IT way of loafing in class: Extending the theory of planned behavior (TPB) to understand students' cyberslacking intentions. *Computers in Human Behavior, 101*, 114-123. https://doi.org/10.1016/j.chb.2019.07.022

Reese, R. M. (2015). *Natural language processing with java*. Packt Publishing.

Rickards, J. P., & Friedman, F. (1978). The encoding versus the external storage hypothesis in note taking. *Contemporary Educational Psychology, 3*(2), 136-143. https://doi.org/10.1016/0361-476X(78)90020-6

Rouquette, A., Hardouin, J. B., Vanhaesebrouck, A., Sébille, V., & Coste, J. (2019). Differential Item Functioning (DIF) in composite health measurement scale: Recommendations for characterizing DIF with meaningful consequences within the Rasch model framework. *PloS one, 14*(4), e0215073. https://doi.org/10.1371/journal.pone.0215073

Rosenberg-Adler, T., & Weintraub, N. (2020). Keyboarding difficulties: Frequency and characteristics among higher education students with handwriting difficulties. *Learning Disabilities Research & Practice, 35*(2), 82-88. https://doi.org/10.1111/ldrp.12220

Roussel, S., Tricot, A., & Sweller, J. (2022). The advantages of listening to academic content in a second language may be outweighed by disadvantages: A cognitive load theory approach. *British Journal of Educational Psychology, 92*(2), 627-644. https://doi.org/10.1111/bjep.12468

Rosner, Z. A., Elman, J. A., & Shimamura, A. P. (2013). The generation effect: Activating broad neural circuits during memory encoding. *Cortex, 49*(7), 1901-1909. https://doi.org/10.1016/j.cortex.2012.09.009

Rost, M. (2013). *Teaching and researching: Listening*. Routledge. https://doi.org/10.4324/9781315833705

Ryu, S., & Kim, J. H. (2003). Learning the lexicon from raw texts for open-vocabulary Korean word recognition. *Proceedings of the Seventh International Conference on Document Analysis and Recognition*. 202-206. https://doi.org/10.1109/ICDAR.2003.1227659

Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin, 2*(6), 110-114. https://doi.org/10.2307/3002019

Service, E., Simola, M., Metsänheimo, O., & Maury, S. (2002). Bilingual working memory span is affected by language skill. *European Journal of Cognitive Psychology, 14*(3), 383-408. https://doi.org/10.1080/09541440143000140

Shimizu, Y., & Zumbo, B. D. (2005). A logistic regression for differential item functioning primer. *Japan Language Testing Association Journal, 7*, 110–124. https://doi.org/10.20622/jltaj.7.0_110

Siegel, J. (2016). A pedagogic cycle for EFL note-taking. *ELT Journal, 70*(3), 275-286. https://doi.org/10.1093/elt/ccv073

Siegel, J. (2018). Did you take "good" notes?: On methods for evaluating student notetaking performance. *Journal of English for Academic Purposes, 35*, 85-92. https://doi.org/10.1016/j.jeap.2018.07.001

Siegel, J. (2020a). *Developing note-taking skills in a second language: Insights from classroom research*. Routledge.

Siegel, J. (2020b). Appreciating translanguaging in student notes. *ELT Journal, 74*(1), 86-88. https://doi.org/10.1093/elt/ccz052

Siegel, J. (2020c). Effects of notetaking instruction on intermediate and advanced L2 English learners: A quasi-experimental study. *Journal of English for Academic Purposes, 46*, 100868. https://doi.org/10.1016/j.jeap.2020.100868

Siegel, J. (2022). Translanguaging in notes: A Swedish perspective. *ASLA: s skriftserie/ASLA Studies in Applied Linguistics, 29*, 99-118. https://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-205566

Siegel, J. (2023a). Pen and paper or computerized notetaking? L2 English students' views and habits. *Computers and Education Open, 4*, 100120. https://doi.org/10.1016/j.caeo.2022.100120

Siegel, J. (2023b). Translanguaging options for note-taking in EAP and EMI. *ELT Journal, 77*(1), 42-51. https://doi.org/10.1093/elt/ccac027

Siegel, J. (2024). Factors affecting notetaking performance. *International journal of Listening, 38*(2), 118-130. https://doi.org/10.1080/10904018.2022.2059484

Siegel, J., & Kusumoto, Y. (2022). A cross-cultural investigation of L2 notetaking: student habits and perspectives. *Journal of Multilingual and Multicultural Development*, 1-17. https://doi.org/10.1080/01434632.2022.2036168

Slotte, V., & Lonka, K. (1999). Review and process effects of spontaneous note-taking on text comprehension. *Contemporary Educational Psychology*, *24*(1), 1-20. https://doi.org/10.1006/ceps.1998.0980

Smoker, T. J., Murphy, C. E., & Rockwell, A. K. (2009). Comparing Memory for Handwriting versus Typing. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 53*(22), 1744-1747. https://doi.org/10.1177/154193120905302218

Song, M. (2011). Note–taking quality and performance on L2 academic listening. *Language Testing, 29*(1), 67-89. https://doi.org/10.1177/0265532211415379

Stefanou, C., Hoffman, L., & Vielee, N. (2008). Note-taking in the college classroom as evidence of generative learning. *Learning Environments Research, 11*, 1-17. https://doi.org/10.1007/s10984-007-9033-0

Steimle, J., Brdiczka, O., & Mühlhäuser, M. (2009). Collaborative paper-based annotation of lecture slides. *Journal of Educational Technology & Society, 12*(4), 125-137. https://www.jstor.org/stable/jeductechsoci.12.4.125

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction, 4(*4), 295-312. https://doi.org/10.1016/0959-4752(94)90003-5

Thompson, K., Leintz, P., Nevers, B., & Witkowski, S. (2004). The integrative listening model: An approach to teaching and learning listening. *The Journal of General Education, 53*(3-4), 225-246. https://doi.org/10.1002/9781444314908

Thompson, K., Leintz, P., Nevers, B., & Witkowski, S. (2010). The integrative listening model: An approach to teaching and learning listening. In A. D. Wolvin (Ed.), *Listening and human communication in the 21st century* (pp. 266-286). Blackwell Publishing.

Tsai-Fu, T. S. A. I., & Wu, Y. (2010). Effects of note-taking instruction and note-taking languages on college EFL students' listening comprehension. *New Horizons in Education*, 58(1), 120-132. http://www.hkta1934.org.hk/NewHorizon/index2.html

Tsui, A. B., & Fullilove, J. (1998). Bottom-up or top-down processing as a discriminator of L2 listening performance. *Applied Linguistics, 19*(4), 432-451. https://doi.org/10.1093/applin/19.4.432

Teng, H. C. (2011). Exploring note-taking strategies of EFL listenerst. *Procedia-Social and Behavioral Sciences, 15*, 480-484. https://doi.org/10.1016/j.sbspro.2011.03.126

Van den Noort, M. W., Bosch, P., & Hugdahl, K. (2006). Foreign language proficiency and working memory capacity. *European Psychologist, 11*(4), 289-296. https://doi.org/10.1027/1016-9040.11.4.289

Vandergrift, L. (2006). Second language listening: Listening ability or language proficiency?. *The Modern Language Journal, 90*(1), 6-18. https://doi.org/10.1111/j.1540-4781.2006.00381.x

Vatcheva, K., Lee, M., McCormick, J. B., & Rahbar, M. H. (2016). Multicollinearity in regression analyses conducted in epidemiologic studies. *Epidemiology, 6,* 227. https://doi.org/10.4172/2161-1165.1000227

Voyer, D., Ronis, S. T., & Byers, N. (2022). The effect of notetaking method on academic performance: A systematic review and meta-analysis. *Contemporary Educational Psychology, 68*, 102025. https://doi.org/10.1016/j.cedpsych.2021.102025

Wagner, E. (2016). Authentic texts in the assessment of L2 listening ability. In J. Banerjee & D. Tsagari (Eds.), *Contemporary second language assessment* (pp.103-123). Bloomsbury Academic. https://doi.org/10.5040/9781474295055

Waite, B. M., Lindberg, R., Ernst, B., Bowman, L. L., & Levine, L. E. (2018). Off-task multitasking, note-taking and lower-and higher-order classroom learning. *Computers & Education, 120*, 98-111. https://doi.org/10.1016/j.compedu.2018.01.007

Wang, X. (2021). An analysis of note-taking strategies: The effect of translanguaging on content comprehension and knowledge retention. *Journal of Language Teaching, 1*(3), 1-20. https://doi.org/10.54475/jlt.2021.020

Welch, B. L. (1947). The generalization of 'student's'problem when several different population varlances are involved. *Biometrika*, *34*(1-2), 28-35. https://doi.org/10.1093/biomet/34.1-2.28

Wilberschied, L. (1998). *The relationships among a variation of immediate recall tasks and measures of L1 writing, L2 achievement, and cognitive strategy use in students of high school Spanish*. [Unpublished doctoral dissertation]. The Ohio State University.

Williams, R. L., & Eggert, A. C. (2002). Notetaking in college classes: Student patterns and instructional strategies. *The Journal of General Education*, *51*(3), 173-199. https://www.jstor.org/stable/27797918

Winter, B. (2020). *Statistics for linguists: An introduction using R*. Routledge.

Witherby, A. E., & Tauber, S. K. (2019). The current status of students' note-taking: Why and how do students take notes?. *Journal of Applied Research in Memory and Cognition, 8*(2), 139-153. https://doi.org/10.1016/j.jarmac.2019.04.002

Wong, S. S. H., & Lim, S. W. H. (2023). Take notes, not photos: Mind-wandering mediates the impact of note-taking strategies on video-recorded lecture learning performance. *Journal of Experimental Psychology: Applied, 29*(1), 124. https://doi.org/10.1037/xap0000375

Worthington, D. L., & Bodie, G. D. (2017). *The sourcebook of listening research: Methodology and measures*. Wiley Blackwell.

Worthington, D. L., & Fitch-Hauser, M. E. (2016). *Listening: Processes, functions, and competency*. Routledge.

Yeager, R., Park, G., & Liao, R. J. (2024). Notetaking as validity evidence: A mixed-methods investigation of question preview in EAP listening assessment. *Journal of English for Academic Purposes, 68*, 101346. https://doi.org/10.1016/j.jeap.2024.101346

Yen, W. M. (1984). Effects of local item dependence on the fit and equating performance of the three-parameter logistic model. *Applied Psychological Measurement, 8*(2), 125-145. https://doi.org/10.1177/014662168400800201

Yue, Z., & Jansem, A. (2022). An Empirical Study of Translanguaging Practice in Interpreters' Note-Taking. *Journal of Language Teaching and Research, 13*(2), 308-314. https://doi.org/10.17507/jltr.1302.10

Zhang, Y. (2008). *Repeater analyses for TOEFL iBT*. ETS. https://www.ets.org/Modes/Research/pdf/RM-08-05.pdf

Zhou, X., Chen, X., & Wang, Z. (2022). The effect of linguistic choices in note-taking on academic listening performance: a pedagogical translanguaging perspective. *International Review of Applied Linguistics in Language Teaching, 61*(2), 775-806. https://doi.org/10.1515/iral-2022-0127

Zohrabi, M., & Esfandyari, F. (2014). The impact of note taking on the improvement of listening comprehension of Iranian EFL learners. *International Journal of English Language and Literature Studies, 3*(2), 165-175. https://ideas.repec.org/a/asi/ijells/v3y2014i2p165-175id661.html