Title of Project:
Video-Mediated Listening Passages and Typed Notetaking: Investigating Their Impact on Comprehension, Test Structure, and Item Performance

Researcher:
Justin Cubilo
University of Hawaii at Manoa
cubiloju@gmail.com

Research Supervisor:
James Dean Brown
University of Hawaii at Manoa
brownj@hawaii.edu

Justin Cubilo

Project Summary

Motivation for the Research

Technology has created many implications for second language (L2) listening assessment, particularly as it relates to the role of visuals and typed note-taking. However, while previous studies have investigated the effects of visuals and typed note-taking on listening test performance, the results of these studies have been contradictory at best, with research indicating that visuals and typed note-taking both help (Kim, Turner, & Perez-Quinones, 2009; Wagner, 2010) and hinder (Muller & Oppenheimer, 2014; Suvorov, 2009) performance. Additionally, while many of these studies have tried to investigate the role of visuals and note-taking on overall test performance, few studies have attempted to investigate the effect that these variables have on item characteristics or on different comprehension subskills. Therefore, the present study attempts to address this gap in the research by investigating the effects of these variables by on item performance and examinee performance in different comprehension skill areas.

Research Questions

1. To what degree do participants' performances on listening tests vary when presented with listening input (i.e., video-based versus audio-only material) and note-taking (i.e., handwritten versus typed) conditions? To what degree do these conditions interact with each other?
2. How do item characteristics differ between the video-based and audio-only conditions? How do they differ between handwritten and typed note-taking conditions?
3. What is the extent to which visual support and note-taking conditions influence examinee's abilities to answer items testing them on different listening comprehension skills?

4. What perceptions and opinions do examinees have of the different conditions to which they are exposed and how do these perceptions and opinions and explanations shed light on the results from the previous questions?

Research Methodology

I employed a quantitative-dominant mixed methods design (Brown, 2014), which incorporated data from a test of academic listening developed for this study as well as qualitative data from a post-test questionnaire that participants were asked to complete. Two hundred participants were asked to take two forms of a listening comprehension task that was designed to measure an academic listening construct focused on comprehension of introductory level university course content. Each test form consisted of three lectures that targeted content from the humanities, social sciences, and natural sciences, and each lecture was followed by 10 multiple-choice questions. Each question set targeted four different subskills representative of the academic domain (identifying the main idea, identifying details, making inferences, and determining speaker attitude/bias). Participants were randomly placed into one of eight experimental groups in which one form of the test was presented in video-mediated format, which included listener access to lecturer gestures and PowerPoint slides, and the other was presented in an audio-only listening format. Likewise, one form required only handwritten note-taking while the other required only typed note-taking. The test was piloted prior to experimental use and was determined to have acceptable reliability, item facility, and item discrimination values, with only three of the thirty items showing need for revision. Examinees finished both forms and then proceeded to complete the post-test questionnaire, which asked them to state their preference for visual and note-taking conditions and to provide reasons for their preference. They also answered questions asking what they found themselves focusing on most in the videos.

Test data were subjected to an ANOVA to determine the overall effect of visual and note-taking on test scores while Rasch analysis and path analysis were used to investigate their effects on item performance and comprehension subskill performance, respectively. Survey data were analyzed inductively to determine codes and themes associated with each response unit and these themes and their associated responses were later used to further explain results obtained from the quantitative analysis.

Summary of Findings

The ANOVA results showed that there was no significant effect of video or typed note-taking on overall test performance. However, low statistical power indicates that a larger sample size may reveal different results in respect to this analysis. Rasch path analysis showed item performance and examinee comprehension subskill performance in different visual and note-taking conditions. The results of the Rasch analysis indicated that items were made slightly easier through the presence of video and typed note-taking. However, this effect did not reach levels of significance. The results of the path analysis indicated that note-taking conditions did not have significant predictive power for examinees' performance in any of the listening comprehension subskills targeted by the items on the listening test. However, it was found that video had significant predictive power on questions that asked examinees to identify details and that performance on these detail-focused questions served as an intermediary variable through which video also had predictive power on examinees' performance on questions asking them to

177 Webster St., # 220, Monterey, CA 93940 USA
Web: www.tirfonline.org / **Email:** info@tirfonline.org

understand main ideas, make inferences, and determine the speaker's attitude. In all cases, this relationship was positive, indicating that the presence of video significantly predicted better performance on different comprehension subskills, most directly the subskill targeting the identification of details.

Qualitative data were also analyzed from the post-test survey using inductive methods to develop a list of thematic categories. Results from this analysis indicated that most participants preferred the video-mediated listening passages because they not only made it easier to understand the material, but also because the visuals from the PowerPoint slides and the lecturer's gestures aided them in getting back on track if they got lost in the middle of the passage, and they made the listening seem more authentic. Those who said they preferred audio-only listening passages primarily gave the reason that the video distracted them and made it difficult to pay attention to the information delivered by the lecturer.

When asked about preferences regarding note-taking conditions, participants overwhelmingly preferred handwritten notes for the primary reason that handwriting made it possible to write notes in their native languages and to show connections between ideas by using circles and arrows. Those who stated that they preferred typing stated that speed and neatness of notes made it the best method for them. Regardless of the preference, most participants stated that they chose their preference based on what they were most comfortable and had the most experience with.

Implications

The findings in this study have a number of implications for the way in which the listening construct is defined. While previous studies have primarily looked at the effect that video-based input has on overall scores for a listening test (Ginther, 2002; Suvorov, 2009; Wagner, 2008), few, if any, have examined the effects it has on different comprehension subskill types. The fact that this study found that visuals significantly contribute to detail comprehension in some situations signals that the impact of visuals may be more refined that previously envisioned. If the listening construct is going to be defined in such a way as to make it so that listening passages are only associated with still pictures of the lecturer or of an object that the lecturer is talking about, then it is necessary to limit this type of test only to subjects that do not see any underlying affects from video-based input. Rather, what seems the safer option is to redefine listening tasks by including the video-based component. In this case, those subjects who do not experience effects from video will produce similar output regardless, while those topics that lead to benefits from its inclusion are allowed to produce these benefits. To remove video-based listening from topics that are affected by it only serves to threaten the construct validity of the test through construct under-representation.

Video-based input also led participants to state a number of other opinions related to the helpfulness of the visuals and the authenticity of the lecture. Even though the slides were constructed to actually contain less information than what one would generally see on slides in a lecture in the real-world domain, participants overwhelmingly stated that they were helpful. In addition, several participants were also clear in stating that they felt that the lectures were more realistic with the video and that it reminded them of attending a real lecture. These findings have several implications. The first of these is that it is clear from the comments, once again, that visuals do play a role in comprehension. Participants made note of both the information on the slides and the lecturer's gestures, body language, and lip movements. This indicates, as previous

177 Webster St., # 220, Monterey, CA 93940 USA
Web: www.tirfonline.org / **Email:** info@tirfonline.org

studies have also done (Sueyoshi & Hardison, 2005), that visual and aural processing are interconnected and that divorcing the two would lead, once again to construct underrepresentation within a listening task. Additionally, comments made by participants related to the visuals providing greater authenticity not only help to provide confirmation that the target language use (TLU) domain is being adequately represented, but also help to establish face validity of the test.

While results related to video-based input seem to provide the greatest number of implications related to the listening construct, note-taking conditions also had an implication worth examining. While the findings for his study indicated that there were no significant differences between scores across note-taking conditions, participants were quite opinionated regarding their actual preferences for note-taking. On the one hand, many participants preferred to take notes by hand, making it so that the current means by which most listening tests require learners to take notes is fair enough for most of the population. On the other hand, however, there were still those who were more accustomed to typing their notes and who stated that they preferred to take notes in this manner because they were better and faster at doing so. This result could call into question the way in which the academic domain is currently conceptualized within academic listening tasks. Even though there were no differences in performance, the fact that many participants stated that one method led to greater comfort than another suggests the need to address this issue. Tests already present themselves as a stress-inducing event. To deprive examinees of the comfort of doing something in a way they would do it in the TLU domain poses potential challenges to the current definition of the TLU domain and, by extension, the construct validity. It may be worth considering whether future test development will take this issue into account and how it can be addressed, though such initiatives may be difficult because they could potentially require that students have constant access to the ability to switch between language typefaces quickly during the test, which may be a difficult feat to accomplish given the number of L1 backgrounds test takers come from.

177 Webster St., # 220, Monterey, CA 93940 USA
Web: www.tirfonline.org / **Email:** info@tirfonline.org

References

Anderson, J. C. (1990). Testing reading comprehension skills. *Reading in a Foreign Language, 6*(2), 425-438.

Anderson, L. W., & Krathwohl, D. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing: A revision of bloom's taxonomy of educational objectives*. New York: Longman.

Anthony, J. J. (2009). Classroom computer experiences that stick: Two lenses on reflective timed essays. *Assessing Writing, 14*, 194-205.

Arnold, J. (2000). Seeing though listening comprehension exam anxiety. *TESOL Quarterly, 34*, 777-786.

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

Ayres, P., & Cierniak, G. (2012). Split attention effect. In N. M. Seel (Ed.), *Encyclopedia of the sciences of learning* (pp. 3172-3175). New York, NY: Springer.

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests.* Oxford, UK: Oxford University.

Bachman, L., & Palmer, A. (2010). *Language assessment in practice.* Oxford: Oxford University.

Baddeley, A. D. (1992). Working memory. *Science, 255*, 556-559.

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

Baltova, I. (1994). The impact of video on comprehension skills of core French students. *Canadian Modern Language Review, 50*(3), 507-531.

Batty, A. O. (2014). A comparison of video- and audio-mediated listening tests with many-facet Rasch modeling and differential distractor functioning. *Language Testing, 32*(1), 3-20.

Bejar, I., Douglas, D., Jamieson, J., Nissan, S., & Turner, J. (2000). *TOEFL 2000 listening framework: A working paper.* Princeton, NJ: Educational Testing Service.

Bloomfield, A., Wayland, S., Rhoades, E., Blodgett, A., Linck, J., & Ross, S. (2010). *What makes listening difficult? Factors affecting second language listening comprehension* (Technical Report No. E.3.1 TTO 81434). College Park, MD: University of Maryland, Center for Advanced Study of Language.

Bodie, G. D., Janusik, L. A., & Valikoski, T.-R. (2008). Priorities of listening research: Four interrelated initiatives. A white paper sponsored by the Research Committee of the International Listening Association. Retrieved from http://www.listen.org/WhitePaper

Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences* (2nd ed.). New York, NY: Routledge.

Brett, P. (1997). A comparative study of the effects of the use of multimedia on listening comprehension. *System, 25*(1), 39-53.

Brown, J. D. (2001a). Point-biserial correlation coefficients. *Shiken: JLT Testing & Evaluation SIG Newsletter, 5*(3), 13-17.

Brown, J. D. (2001b). *Using surveys in language programs.* Cambridge, UK: Cambridge University Press.

Brown, J. D. (2005). *Testing in language programs: A comprehensive guide to English language assessment*. New York, NY: McGraw-Hill.

Brown, J. D. (2014). *Mixed methods research for TESOL*. Edinburgh, UK: Edinburgh University.

Brown, J. D., Trace, J., Janssen, G., & Kozhevnikova, L. (2016). How well do cloze items work and why? In C. Gitsaki & C. Coombe (Eds.), *Current issues in language evaluation, assessment, and testing: Research and Practice* (pp. 2-39). Newcastle upon Tyne, England: Cambridge Scholars Publishing.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136-162). Newbury Park, CA: Sage Publishers.

Buck, G. (2001). *Assessing listening.* Cambridge, UK: Cambridge University Press.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*, 81-105.

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction, 8*, 293-332.

Chapelle, C. A., Enright, M. K., & Jamieson, J. (2010). Does an argument-based approach to validity make a difference? *Educational Measurement: Issues and Practice, 29*(1), 3-13.

Chapelle, C. A., Enright, M. K., & Jamieson, J. M. (2008). Test score interpretation and use. In C. A. Chapelle, M. K. Enright, & J. M. Jamieson (Eds.), *Building a validity argument for the Test of English as a Foreign Language* (pp. 1-25). New York, NY: Routledge.

Chaudron, C., Loschky, L., & Cook, J. (1994). Second language listening comprehension and lecture note-taking. In J. Flowerdew (Ed.), *Academic listening: Research perspectives* (pp. 75-92). Cambridge, UK: Cambridge University Press.

Chung, U. K. (1994). *The effect of audio, a single picture, multiple pictures, or video on second-language listening comprehension*. Unpublished PhD dissertation, University of Illinois at Urbana-Champaign.

Coniam, D. (2001). The use of audio or video comprehension as an assessment instrument in the certification of English language teachers: A case study. *System, 29*, 1-14.

Cronbach, L. J. (1988). Five perspectives on validity argument. In H. Wainer & H. Braun (Eds.), *Test validity* (pp. 3-17). Hillsdale, NJ: Lawrence Erlbaum.

Cubilo, J. & Winke, P. (2013). Redefining the L2 listening construct with an integrated writing task: Considering the impact of visual-cue interpretation and note-taking. *Language Assessment Quarterly, 10*, 371-397.

Cutler, A. (2012). *Native listening: Language experience and the recognition of spoken words*. Cambridge, MA: The MIT Press.

Davidson, F., & Lynch, B. K. (2001). *Testcraft: A teacher's guide to writing and using language test specification.* New Haven, CT: Yale University Press.

Davies, A. (Ed.). (1997). Ethics in language testing. *Language Testing, 14*.

Denzin, N. K. (1978). *The research act: A theoretical introduction to sociological methods.* New York, NY: Praeger.

Desnoyers, L. (2011) Toward a taxonomy of visuals in science communication. *Technical Communication, 58*(2), 119-134.

Dunkel, P., & Davy, S. (1989). The heuristic of lecture notetaking: Perceptions of American and international students regarding the value and practice of notetaking. *English for Specific Purposes, 8*(1), 33-50.

Eckes, T. (2009). Many-facet Rasch measurement. *Reference supplement to the manual for relating language examinations to the Common European Framework of Reference for Languages: Learning, teaching, assessment*. Frankfurt, Germany: Peter Lang.

English, S. L. (1982, May). Kinesics in academic listening. Paper presented at the 16th annual convention of Teachers of English to Speakers of Other Languages, Honolulu, HI. (ERIC Document Reproduction Service No. ED 218 976).

English, S. L. (1985). Kinesics in academic lectures. *The ESP Journal, 4*(2), 161-170.

Field, J. (2008). *Listening in the language classroom*. Cambridge: Cambridge University.

Fink, J. L. (2010). Why we banned use of laptops and "scribe notes" in our classroom. *American Journal of Pharmaceutical Education, 74*(6), Article 114.

Flowerdew, J. (1994). Research of relevance to second language lecture comprehension – An overview. In J. Flowerdew (Ed.), *Academic listening: Research perspectives* (pp. 7-29). Cambridge: Cambridge University.

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). Oxford, UK: Wiley-Blackwell.

Gage, N. L. (1989). The paradigm wars and their aftermath: A "historical" sketch of research on teaching since 1989. *Educational Researcher, 18*(7), 4-10.

Ginther, A. (2002). Context and content visuals and performance on listening comprehension stimuli. *Language Testing, 19*(2), 133-167.

Goodwin, S. J. (2017). *Locus of control in L2 English listening assessment*. Unpublished doctoral dissertation, Georgia State University, Atlanta, GA.

Greene, J. C. (2008). Is mixed methods social inquiry a distinctive methodology? *Journal of Mixed Methods Research, 2*(1), 7-22.

Greene, J. C. (2011). The construct(ion) of validity as argument. In H. T. Chen, S. I. Donaldson, & M. M. Mark (Eds.), *Advancing validity in outcome evaluation: Theory and practice, new directions for evaluation* (pp. 81-92). San Francisco, CA: Jossey-Bass.

Greene, J. C., Caracelli, V. J., & Graham, W. F. (1989). Toward a conceptual framework for mixed-method evaluation designs. *Educational Evaluation and Policy Analysis, 11*(3), 255-274.

Gruba, P. (1993). A comparison study of audio and video in language testing. *JALT Journal, 15*(1), 85-88.

Gruba, P. (1999). *The role of digital video media in second language listening comprehension.* Unpublished PhD dissertation, Department of Linguistics and Applied Linguistics, University of Melbourne. Retrieved December 21, 2016, from http://eprints.unimelb.edu.au/archime/00000244/

Gruba, P. (2006). Playing the videotext: A media literacy perspective on video-mediated L2 listening. *Language Learning & Technology, 10*(2), 77-92.

Guba, E. G., & Lincoln, Y. S. (1989). *Fourth generation evaluation*. Newbury Park, CA: Sage Publications.

Hadar, U., Wenkert-Olenik, D., Krauss, R., & Soroker, N. (1998). Gesture and the processing of speech: Neuropsychological evidence. *Brain and Language, 62*, 107-126.

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

Harsch, C., & Martin, G. (2012). Adapting CEF-descriptors for rating purposes: Validation by a combined rater training and scale revision approach. *Assessing Writing, 17*, 228-250.

Hartley, J. & Davies, I. K. (1978). Note-taking: A critical review. *Innovations in Education & Training International, 15*(3), 207-224.

Hatch, E., & Lazaraton, A. (1991). *The research manual: Design and statistics for applied linguistics.* Boston, MA: Heinle & Heinle.

Hayati, A. M., & Jalilifar, A. R. (2009). The impact of note-taking strategies on listening comprehension of EFL learners. *Canadian English Language Teaching, 2*(1), 101-111.

Horz, H., & Schnotz, W. (2010). Cognitive load in learning with multiple representations. In J. L. Plass, R. Moreno, & R. Brünken (Eds.), *Cognitive load theory* (pp. 229-252). Cambridge, UK: Cambridge University Press.

Hu, L.-T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1-55.

In'nami, Y. (2006). The effects of test anxiety on listening test performance. *System, 34*, 317-340.

Institute of International Education. (2016). "International Student Enrollment Trends, 1948/49-2015/16." *Open Doors Report on International Educational Exchange*. Retrieved from http://www.iie.org/opendoors

International Listening Association (1995). A ILA definition of listening. *The Listening Post, 53*, 1-5.

Jang, E. E. (2005). *A validity narrative: Effects of reading skills diagnosis on teaching and learning in the context of NG TOEFL.* Unpublished PhD dissertation, University of Illinois at Urbana-Champaign.

Jang, E. E. (2009). Cognitive diagnostic assessment of L2 reading comprehension ability: Validity arguments for Fusion Model application to *LanguEdge* assessment. *Language Testing, 26*(1), 31-73.

Jang, E. E., Wagner, M., & Park, G. (2014). Mixed methods research in language testing and assessment. *Annual Review of Applied Linguistics, 34*, 123-153.

Jensen, J. L., McDaniel, M. A., Woodard, S. M., & Kummer, T. A. (2014). Teaching to the test…or testing to teach: Exams requiring higher order thinking skills encourage greater conceptual understanding. *Educational Psychology Review, 26*(2), 307-329.

Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness.* Cambridge, UK: Cambridge University Press.

Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher, 33*(7), 14-26.

Johnson, R. B., Onwuegbuzie, A. J., & Turner, L. A. (2007). Toward a definition of mixed methods research. *Journal of Mixed Methods Research, 1*(2), 112-133.

Kane, M. T. (1992). An argument-based approach to validation. *Psychological Bulletin, 112*, 527-535.

Kane, M. T. (2001). Current concerns in validity theory. *Journal of Educational Measurement, 38*, 319-342.

Kane, M. T. (2002). Validating high-stakes testing programs. *Educational Measurement: Issues and Practice, 18*, 5-17.

Kane, M. T. (2006) *Validation*. In R. Brennan (Ed.), *Educational Measurement*, 4th ed. (pp. 17-64), Westport, CT: American Council on Education and Praeger.

Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement, 50*, 1-73.

Kane, M. T., Crooks, T. J., & Cohen, A. S. (1999). Validating measures of performance. *Educational Measurement: Issues and Practice, 18*, 5-17.

Kellerman, S. (1992). "I see what you mean": The role of kinesic behavior in listening and implications for foreign and second language learning. *Applied Linguistics, 13*(3), 239-258.

Kim, K., Turner, S. A., & Perez-Quinones, M. A. (2009). Requirements for electronic note-taking systems: A field study of note-taking in university classrooms. *Education and Information Technologies, 14*(3), 255-283.

King, P. E., & Behnke, R. R. (1989). The effect of time-compressed speech on comprehensive, interpretive, and short-term listening. *Human Communication Research, 15*(3), 428-443.

Kintsch, W. (1998). *Comprehension.* Cambridge, UK: Cambridge University Press.

Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). New York, NY: Guilford Press.

Kline, R. B. (2012). Assumptions in structural equation modeling. In R. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 111-125). New York, NY: Guilford Press.

Ladas, H. S. (1980). Note-taking on lectures: An information-processing approach. *Educational Psychologist, 15*, 44-53.

Lado, R. (1961). *Language testing: The construction and use of language tests.* London, UK: Longman.

Lee, H., & Winke, P. (2013). The differences among three-, four-, and five-option-item formats in the context of a high-stakes English-language listening test. *Language Testing, 30*(1), 99-123.

Leech, N. L., & Onwuegbuzie, A. J. (2009). A typology of mixed methods research designs. *Quality & Quantity, 43*(2), 265-275.

Levelt, W. J. M. (1993). Language use in normal speakers and its disorders. In G. Blanken, J. Dittmann, H. Grimm, J. C. Marshall, & C.-W. Wallesch (Eds.), *Linguistic disorders and pathologies* (pp. 1-15). Berlin, Germany: De Gruyter.

Linacre, J. M. (2014). Facets computer program for many-facet Rasch measurement, version 3.71.4. Beaverton, Oregon: Winsteps.

Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological Reports, Monograph Supplement, 3*, 635-694.

Londe, Z. C. (2009). The effects of video media in English as a second language listening comprehension tests. *Issues in Applied Linguistics, 17*(1), 41-50.

Lund, R. J. (1990). A taxonomy for teaching second language listening. *Foreign Language Annals, 23*(2), 105-115.

Lynch, T. (2011). Academic listening in the 21st century: Reviewing a decade of research. *Journal of English for Academic Purposes, 10*(2), 79-88.

Mayer, R. E. (Ed.). (2005). *The Cambridge handbook of multimedia learning.* Cambridge, UK: Cambridge University Press.

McCuistion, P. J. (1991). Static vs. dynamic visuals in computer-assisted instruction. *Engineering Design Graphics Journal, 55*(2), 25-33.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264*, 746-748.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (pp. 13-103). New York, NY: Macmillan.

Mislevy, R., Steinberg, L., & Almond, R. (2003). On the structure of educational assessments. *Measurement: Interdisciplinary Research and Perspectives, 1,* 3-62.

Moreno, R., & Park, B. (2010). Cognitive load theory: Historical development and relation to other theories. In J. L. Plass, R. Moreno, & R. Brünken (Eds.), *Cognitive load theory* (pp. 9-28). Cambridge, UK: Cambridge University Press.

Morrel Samuels, P., Krauss, R. M. (1992). Word familiarity predicts temporal asynchrony of hand gestures and speech. *Journal of Experimental Psychology: Learning, Memory and Cognition, 18,* 615-662.

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.

Muthén, L. K., & Muthén, B. O. (2014). *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.

Nagel, T. (1986). *The view from nowhere.* Oxford, UK: Oxford University Press.

Nagle, S. J., & Sanders, S. L. (1986). Comprehension theory and second language pedagogy. *TESOL Quarterly, 20*(1), 9-26.

NVivo qualitative data analysis software (Version 10) [Computer software]. (2014). Doncaster, Australia: QSR International Pty Ltd.

Ochs, E., & Schieffelin, B. (2009). Language acquisition and socialization: Three developmental stories and their implications. In A. Duranti (Ed.), *Linguistic anthropology: A reader* (2nd ed.) (pp. 296-328). Malden, MA: Wiley-Blackwell.

Ockey, G. J. (2007). Construct implications of including still image or video in computer-based listening tests. *Language Testing, 24*(4), 517-537.

Ockey, G. J., Papageorgiou, S., & French, R. (2016). Effects of strength of accent on an L2 interactive lecture listening comprehension test. *International Journal of Listening, 30*(1-2), 84-98.

Olson, K. (2003). LSAT listening assessment: Theoretical background and specifications. *Law School Admission Council (LSAC) Research Report 03-02.* Retrieved from http://www.lsac.org/lsacresources/Research/rr/pdf/RR-03-02.pdf

Onwuegbuzie, A. J., & Johnson, R. B. (2006). The validity issue in mixed research. *Research in the Schools, 13,* 48-63.

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.

Pettersson, R. (2002). *Information design: An introduction.* Amsterdam: John Benjamins Publishing Company.

Peverly, S. T., Garner, J. K., & Vekaria, P. C. (2013). Both handwriting speed and selective attention are important to lecture note-taking. *Reading and Writing: An Interdisciplinary Journal, 27*, 1-30.

Piolat, A., Olive, T., & Kellogg, R. T. (2005). Cognitive effort during note taking. *Applied Cognitive Psychology, 19*, 291-312.

Plonsky, L., & Derrick, D. J. (2016). A meta-analysis of reliability coefficients in second language research. *The Modern Language Journal, 100*(2), 538-553.

Progrosh, D. (1996). Using video for listening assessment: Opinions of test-takers. *TESL Canada Journal, 14*, 34-44.

Purdy, M. W. (2010). Qualitative research: Critical for understanding listening. In A. D. Wolvin (Ed.), *Listening and human communication in the 21st century* (pp. 33-45). Oxford, UK: Wiley-Blackwell.

Rost, M. (2011). *Teaching and researching listening* (2nd ed.). Harlow, UK: Pearson.

Rubin, J. (1995). The contribution of video to the development of competence in listening. In D. Mendelsohn, & J. Rubin (Eds.), *A guide for the teaching of second language listening* (pp. 151-165). San Diego, CA: Dominie Press.

Schnotz, W. (2005) An integrated model of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 49-69). Cambridge, UK: Cambridge University Press.

Smidt, E., & Hegelheimer, V. (2004). Effects of online academic lectures on ESL listening comprehension, incidental vocabulary acquisition, and strategy use. *Computer Assisted Language Learning, 17*(5), 517-556.

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 – 2009, 53*, 1744-1747.

Song, M.-Y. (2012). Note-taking quality and performance on an L2 academic listening test. *Language Testing, 29*(1), 67-89.

Stacy, E. M., & Cain, J. (2015). Note-taking and handouts in the digital age. *American Journal of Pharmaceutical Education, 79*(7), 1-6.

Sueyoshi, A., & Hardison, D. M. (2005). The role of gestures and facial cues in second language listening comprehension. *Language Learning, 55*, 661-699.

Suvorov, R. (2008). *Context visuals in L2 listening tests: The effectiveness of photographs and video vs. audio-only format*. Unpublished master's thesis, Iowa State University, Ames, IA.

Suvorov, R. (2009). Context visuals in L2 listening tests: The effects of photographs and video vs. audio-only format. In C. A. Chapelle, H. G. Jun, & I. Katz (Eds.) *Developing and evaluating language learning materials* (pp. 53-68). Ames, IA: Iowa State University.

Suvorov, R. (2013). Interacting with visuals in L2 listening tests: An eye-tracking study. Unpublished doctoral dissertation, Iowa State University, Ames, IA.

Suvorov, R. (2015). The use of eye tracking in research on video-based second language L2 listening assessment: A comparison of context videos and content videos. *Language Testing, 32*(4), 463-483.

Sweller, J., & Chandler, P. (1994). Why some material is difficult to learn. *Cognition and Instruction, 12*(3), 185-233.

Teddlie, C., & Tashakkori, A. (2006). A general typology of research designs featuring mixed methods. *Research in the Schools, 13*(1), 12-28.

Teng, H. (2011). Exploring note-taking strategies of EFL listeners. *Procedia – Social and Behavioral Sciences, 15*, 480-484.

Toulmin, S. E. (1958). *The uses of argument.* Cambridge, UK: Cambridge University Press.

Toulmin, S. E. (2003). *The uses of argument: Updated edition.* Cambridge, UK: Cambridge University Press.

Vandergrift, L. (2010). Researching listening. In B. Paltridge & A. Phakiti (Eds.), *Continuum companion to research methods in applied linguistics* (pp. 160-173). London, UK: Continuum International Publishing Group.

Vandergrift, L., & Goh, C. C. M. (2011). *Teaching and learning second language listening: Metacognition in action.* New York, NY: Routledge.

Vanderplank, R. (2010). Déjà vu? A decade of research on language laboratories, television and video in language learning. *Language Teaching, 43*(1), 1-37.

von Raffler-Engel, W. (1980). Kinesics and paralinguistics: A neglected factor in second-language research and teaching. *Canadian Modern Language Review, 36*(2), 225-237.

Voss, J. L., Gonsalves, B. D., Federmeier, K. D., Tranel, D., & Cohen, N. J. (2011). Hippocampal brain-network coordination during volitional exploratory behavior enhances learning. *Nature Neuroscience, 14*(1), 115-120.

Wagner, E. (2002). Video listening tests: A pilot study. *Working Papers in TESOL & Applied Linguistics, Teachers College, Columbia University, 2*(1). Retrieved from http://journals.tc-library.org/index.php/tesol/issue/view/2

Wagner, E. (2006). *Utilizing the visual channel: An investigation of the use of video texts on tests of second language listening ability.* Unpublished doctoral dissertation, Teachers College, Columbia University, New York.

Wagner, E. (2007). Are they watching? Test-taker viewing behavior during an L2 video listening test. *Language Learning & Technology, 11*(1), 67-86.

Wagner, E. (2008). Video listening tests: What are they measuring? *Language Assessment Quarterly, 5*(3), 218-243.

Wagner, E. (2010a). Test-takers' interaction with an L2 video listening test. *System, 38*, 280-291.

177 Webster St., # 220, Monterey, CA 93940 USA
Web: www.tirfonline.org / **Email:** info@tirfonline.org

Wagner, E. (2010b). The effect of the use of video texts on ESL listening test-taker performance. *Language Testing, 27*, 493-513.

Wagner, E. (2013). An investigation of how the channel of input and access to test questions affect L2 listening test performance. *Language Assessment Quarterly, 10*(2), 178-195.

Waszak, C., & Sines, M. (2003). Mixed methods in psychological research. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods in social and behavioral research* (pp. 557-576). Thousand Oaks, CA: Sage Publishers.

Webb, E. J., Campbell, D. T., Schwartz, R. D., & Sechrest, L. (1966). *Unobtrusive measures: Nonreactive research*. Chicago, IL: Rand McNally.

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

Wright, B. D., & Linacre, J. M. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions, 8*, 370.

177 Webster St., # 220, Monterey, CA 93940 USA
Web: www.tirfonline.org / **Email:** info@tirfonline.org