Isbell, D. R. (2020). Diagnostic language assessment for L2 pronunciation: A worked example. In O. Kang, S. Staples, K. Yaw, & K. Hirschi (Eds.), *Proceedings of the 11th Pronunciation in Second Language Learning and Teaching conference*, ISSN 2380-9566, Northern Arizona University, September 2019 (pp. 127–140). Ames, IA: Iowa State University.

DIAGNOSTIC LANGUAGE ASSESSMENT FOR L2 PRONUNCIATION: A WORKED EXAMPLE

Daniel R. Isbell, University of Hawai‘i at Mānoa

Although pronunciation teaching has come back into fashion, pronunciation assessment remains mostly focused on evaluating global qualities (Isaacs & Trofimovich, 2017; Kang & Ginther, 2017). This assessment approach offers little to learners and teachers who desire finer-grained information about their pronunciation abilities to help guide their learning activities. Diagnostic language assessment, however, can be used to identify learner strengths and weaknesses to guide learning. While diagnostic assessment holds great promises for informing learning, there are relatively few examples of truly diagnostic tests, and this paucity is especially stark for pronunciation. This article describes the design of a diagnostic pronunciation test for Korean as a second language, which includes tasks targeting both production and perception of Korean phonemes. From this test, learners receive information on how well they perceive and produce Korean sounds, which can raise awareness of difficulties and steer learners toward suitable learning activities. In addition to describing test and task specifications, key principles for pronunciation diagnosis are reviewed. Finally, an approach to interpreting and validating diagnostic test scores is discussed.

INTRODUCTION

While second language (L2) pronunciation instruction is back in vogue and perhaps even enjoying a renaissance (Levis, 2019), pronunciation assessment, in particular, has not evolved to serve the purposes of teaching and learning. Most contemporary work on pronunciation assessment falls under the umbrella of *proficiency assessment*, which focuses primarily on evaluating global qualities of pronunciation skills, often in the context of assessing pronunciation as a subcomponent of speaking skills (Isaacs & Trofimovich, 2017; Kang & Ginther; 2017). Such assessments are certainly useful, but they do little to address the questions most relevant to teaching individual learners, like “What specific pronunciation difficulties does this student actually have?” Unfortunately, types of assessments more relevant to pronunciation learning and teaching receive little attention, and in some ways, have not shown much development from Lado’s (1961) recommendations (e.g., assessing individual L2 segments, using pictures to elicit sounds), many of which, to be fair, are still useful. This lack of development is especially unfortunate given experts’ calls to diagnose learner needs and individualize instruction (e.g., Levis, 2018).

One promising avenue for assessment that is more attuned to pronunciation teaching and learning is diagnostic language assessment (DLA; Alderson, 2005; Alderson et al., 2015). In DLA, the focus is on identifying the specific strengths and weaknesses of individual learners rather than placing them on a continuum of general ability, as is the case with typical proficiency assessment. The overarching goal in DLA is to promote subsequent learning: By identifying student weaknesses, teachers and learners know what to prioritize in their teaching and learning activities. Whereas proficiency assessments (and most forms of language assessment) are designed in reference to a description of a target ability (e.g., “speaking”), DLAs go beyond this and

incorporate what is known about language development into the design of assessments. In other words, DLA reveals information about the targeted skill/knowledge as well as provides insights on developmental progress. DLA is well-positioned to meet the recommendations of experts in pedagogy, such as Levis (2018, p. 239), who proposes that "teachers will take note of each student's challenges either through diagnosing the student's speech or through noticing difficulties during class."

When diagnosing learners' pronunciation difficulties, intelligibility is the most suitable criterion for evaluation in instructionally-relevant assessment: Intelligibility is the primary goal of communication, and intelligible pronunciation, rather than native-like, is a goal within reach of L2 learners and thereby aimed for in instruction (Levis, 2005). It is now intelligibility, rather than nativeness, that constitutes descriptions of pronunciation ability in proficiency frameworks like the Common European Framework of Reference (CEFR, Council of Europe, 2017), with L2 learners attaining more intelligible pronunciation as their global proficiency increases. While such a trend is likely to exist across learners, individual deviation from that trend is likely. For example, some learners at lower levels of overall proficiency may exhibit highly intelligible pronunciation, and some otherwise high-proficiency learners may still struggle in terms of pronunciation. From a diagnostic perspective, these individual differences are of primary concern.

In this paper, I exemplify DLA through a test I developed, the Korean Pronunciation Diagnostic (KPD), that supports learners in developing more intelligible pronunciation. In what follows, I discuss some key principles for pronunciation diagnosis, describe the design of the KPD, and lay out an approach to investigating its validity. This latter part is presented in brief due to space limitations but included due to its importance in assessment: Making a test is only the beginning of assessing learners, and valid use of test scores should always be kept in mind.

Some Key Principles for Diagnosing Pronunciation

It is desirable for a pronunciation diagnostic to yield information about strengths and weaknesses in word-level pronunciation phenomena (e.g., segments, lexical stress; see Levis, 2018). While discourse-level phenomena, such as suprasegmentals, do have an impact on intelligibility, empirical research on intelligibility suggests that segmental features have considerable effects, too (Kang et al., 2018; Loewen & Isbell, 2017). Thus, at least as a starting point, a focus on segmental pronunciation features in diagnosis is justifiable. In the case of Korean, an initial focus on segmental features is further supported by its phonological characteristics. Korean is a syllable-timed language without lexical stress (e.g., English), lexical tone (e.g., Chinese), or other word-level suprasegmental features used to distinguish lexical items, resulting in functional loads entirely dependent on consonant and vowel relations (Oh et al., 2015). At the same time, Korean's phoneme inventory is average in size and its syllable structure is fairly restrictive, resulting in considerable neighborhood density in the Korean lexicon (i.e., number of words differentiated by a single segment; Holliday et al., 2017). In other words, intelligible production of Korean segments is especially crucial for a speaker's intended words to be accurately identified by an interlocutor.

In terms of pronunciation learning, the roles of perception of speech sounds and awareness of one's own pronunciation are particularly relevant to DLA. A key finding of pronunciation development is the link between perception and production: Accuracy in production is linked to accuracy in

perception, and there is evidence that improvements to perception can lead to improvements in production (Flege, 1991; Lee & Lyster, 2017; Sakai & Moorman, 2018; Thomson, 2011). At a more general level, learner awareness of sounds and their own accuracy and directing attention to acoustic form and articulation in instruction has also been found to promote pronunciation development (Guion & Pedersen, 2007; Moyer, 2014; Saito, 2019). This suggests that making students aware of their strengths and weaknesses is an important goal of diagnosis and implies that diagnostic information should be readily interpretable for learners.

One challenge in assessing pronunciation, especially specific aspects of it, is avoiding influence from factors unrelated to pronunciation abilities. For instance, Crowther et al. (2015) found that more difficult speaking tasks resulted in poorer listener-based judgments of global pronunciation qualities. Other research has shown that (malformed) lexical knowledge can result in non-systematic pronunciation or perception errors (e.g., Amengual, 2016) and that pronunciation performance while reading text aloud may not faithfully reflect general competence (Levis & Barriuso, 2012). There are two strategies a test developer might consider to address these issues: (1) designing tasks in ways that limit undesirable influences, and (2) including multiple tasks to counter-balance or mitigate the drawbacks of any one task. With respect to the former strategy, consulting psycholinguistic models of language processing is instructive (Harding et al., 2015). John Field's process models of L2 speech production and perception (2011, 2013) suggest that phonological knowledge and phonetic encoding/acoustic decoding might be isolated (to the degree possible) in tasks which largely exclude higher-level conceptualization/comprehension, syntactic encoding, and lexical search.

Finally, practicality demands consideration. DLA is low-stakes, and resource barriers, such as cost, time, technology, or advanced expertise, are likely to preclude it from being carried out. Thus, developing low-cost (or free) tools that can be administered and scored in little time by any teacher or tutor is desirable for a pronunciation diagnostic.

DESIGN OF THE KPD

Test Purpose

The purpose of the KPD is to yield detailed information about strengths and weaknesses in Korean segmental production and perception for learners of any L1 background and proficiency level. The diagnostic information is intended to motivate and inform decisions in subsequent pronunciation learning activities, making it appropriate to use with individual learners who wish to develop their pronunciation abilities. For instance, a learner may approach a teacher and ask for guidance in improving their pronunciation abilities. The teacher could then use the KPD to identify the learner's priorities for study. The KPD was designed with practicality in mind: The test is freely available and requires only a recorder, about 15 minutes for administration, and perhaps another 15 minutes to score and process results. This practicality supports the low-stakes, learning-focused purpose of using the KPD. The KPD is not to be used in high-stakes decision making or for assessing overall pronunciation quality.

Test Specifications

The KPD has two parts, Production and Perception, with each part containing two tasks (Table 1). The task selections and designs were informed by recommendations from Derwing and Munro (2015), Harding et al. (2015), and Lado (1961) pertaining to isolating specific targets, obtaining multiple observations of targets, avoiding potential interference from higher-level processing demands, and counterbalancing the advantages and disadvantages of any individual task.

In each modality, there are several items targeting each Korean phoneme; the exact number per phoneme varies according to (a) its allophonic distribution and permissible syllable contexts (e.g., /k/ has three allophones and may appear in a syllable onset or coda, as well as intervocalically) and, relevant to production tasks, (b) its prevalence in common words (e.g., /a, i, o/ are found in many common words). At least 4 items for each phoneme are included across production tasks, and at least 3 are included across perception tasks. The rationale for having multiple items for each phoneme is that sounds which are more frequently mispronounced are likely to have greater negative effects on a speaker's intelligibility (Levis, 2018).

One task in each section features picture stimuli, while the other task features written stimuli. These two types of tasks were included in order to counter-balance the undesirable influences of poor vocabulary knowledge and poor sound-symbol correspondence knowledge, respectively. Sum scores for each phoneme in each modality are computed and reported. Supplementary information, such as challenging articulatory features across several phonemes, can also be included in the score report but is not the primary source of diagnostic information. Information on scoring each task and score reporting follow. To access the full test materials, readers may visit https://sites.google.com/view/koreanpronunciationdiagnostic.

Table 1
KPD Design Summary

Section	Task	Brief Item Specification	Number of Items
Production	Picture Naming	Item: picture of a concrete noun Response: speaking the matching word	154 (in 35 words)
	Nonword Reading	Item: 1-2 syllable nonword Response: reading aloud the nonword	63
Perception	Pronunciation Judgment	Item: picture of a concrete noun + audio recording of the word Response: forced choice whether audio recording was (in)correct	72 (plus 40 filler items)
	Nonword Identification	Item: audio recording of a 1-2 syllable nonword Response: forced choice between two written 1-2 syllable nonwords	63

Task Designs

Production Task 1: Picture Naming

In the Picture Naming tasks, learners are required to say the word which corresponds to a picture they are shown (Figure 1). This is a task format commonly used in L1 speech development research and clinical practice that taps into learners' knowledge of phonological form-meaning mappings, followed by activation of encoding and articulation processes. One potential drawback of this task is poor lexical knowledge. This was mitigated by generally selecting target words within the first 1,500 most common Korean words (Lee et al., 2017) and other words known to be common in Korean textbooks. Pictures were primarily drawn from databases used in psycholinguistic research (Brodeur et al., 2010; Duñabeitia et al., 2017). Despite the use of highly frequent target words and iconic pictures, some learners would occasionally offer a non-target word or fail to know the word at all. In these cases, the tester may prompt the learner with a hint, or if that fails, may say the word for the learner to repeat. This task is subjectively scored, ideally by an experienced teacher or tutor. For each word, the scorer listens to the learner's recorded response and crosses out any phonemes that were not readily identifiable. Crossed-out sounds are awarded a value of 0, while acceptable phonemes are assigned a value of

1.

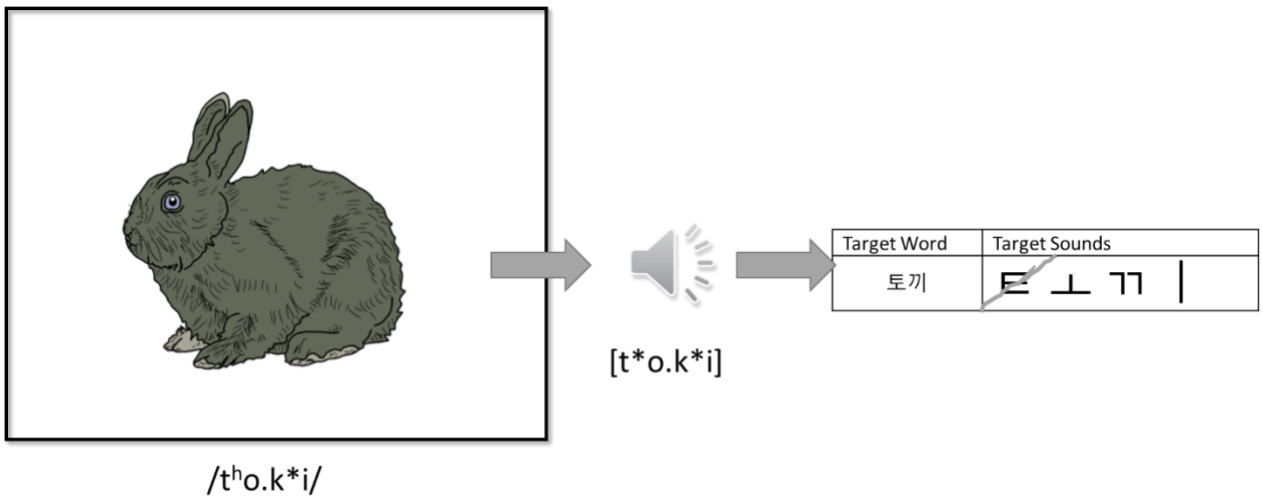

Figure 1. Schematic depicting the Picture Naming task. In this example, the learner's articulation of *rabbit* was not entirely intelligible due to the articulation of the first phoneme ([t*] instead of [t^h]).

Production Task 2: Nonword Reading

The Nonword Reading task requires learners to read a 1-2 syllable nonword aloud, with each nonword only being evaluated for one target sound (Figure 2). The nonwords take the form of (C/G)V, VC, or VCV. Here, written letters are used to tap into phonological knowledge (symbol-sound correspondence), which then leads to phonological encoding and articulation. This task, while potentially contaminable by poor orthographic(-phonological) knowledge, counter-balanced

the Picture Naming task's dependence on lexical knowledge and potential pronunciation difficulties related to co-articulation effects. An advantage of this task was the flexibility to include phonemes in certain contexts that were difficult to target in high-frequency, imageable real words. To mitigate poor orthographic knowledge, nonwords are constructed with maximal transparency in terms of symbol-sound correspondences. Nonword Reading, like Picture Naming, is subjectively scored. However, the scorer only judges whether the target sound was readily identifiable; the rest of the nonword is ignored. Unintelligible target sounds are coded as 0, while intelligible sounds are assigned a score of 1.

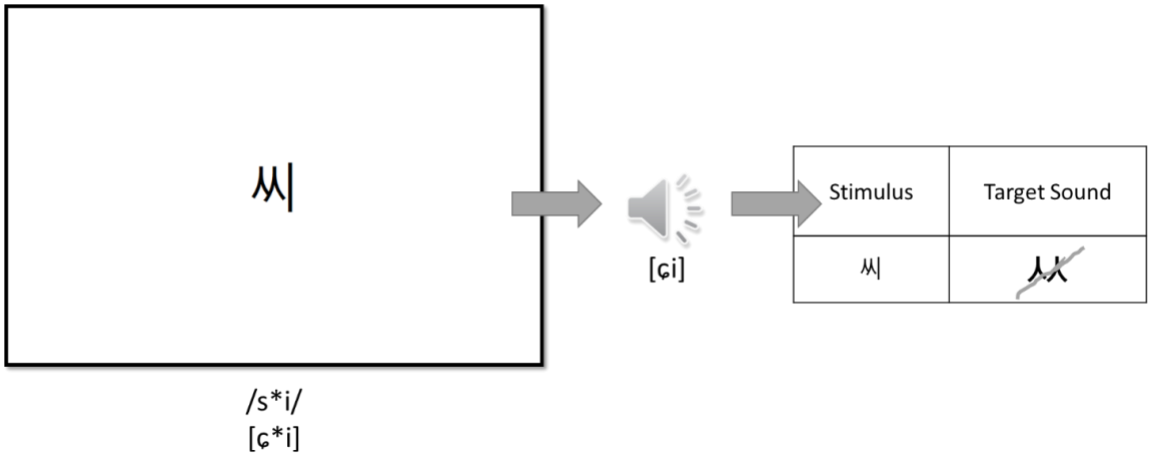

Figure 2. Schematic depicting the Nonword Reading task. In this example, the learner's production of /s*/ would be judged as incorrect (not readily intelligible).

Perception Task 1: Pronunciation Judgment

In the Pronunciation Judgment task, learners are shown a picture of a common word and must judge whether an audio recording of that word is articulated correctly (Figure 3). The audio recordings, all created by a female speaker of the widely familiar Seoul dialect, are either entirely acceptable or intentionally articulated with one noticeable deviation (the noticeability was verified through piloting with seven native speakers of Korean who were primary speakers of or otherwise highly familiar with Seoul Korean). Deviations are substitutions of one phoneme with another phoneme that shares some articulatory settings (e.g., /k/ substituted by /k*/). This type of task has been fruitfully deployed in psycholinguistics research to probe receptive phonological knowledge (e.g., Amengual, 2016). Like the Picture Naming task, responses to this task can potentially be contaminated by poor lexical knowledge, but in this case such concern is slightly mitigated by being able to hear an articulation of the target word, which can aid in activating phonological form knowledge. Despite the possibility of response contamination due to poor lexical knowledge, this task requires no knowledge of orthographic form or symbol-sound correspondence. This task is scored objectively; student responses are compared against an answer key. Correct answers earn a score of 1, while incorrect answers are awarded 0.

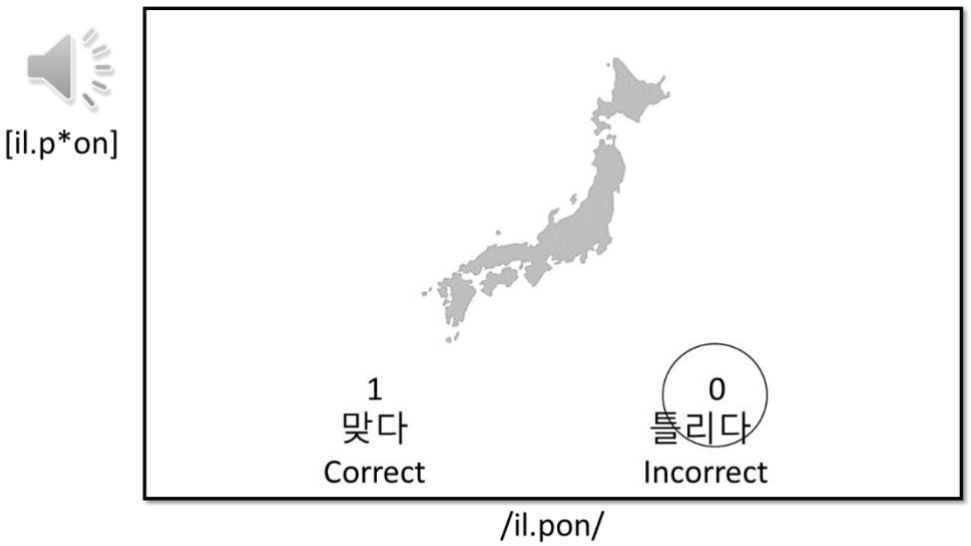

Figure 3. Schematic depicting the Pronunciation Judgment task. In this example, learners hear a recording of *Japan* (/il.pon/) that substituted /p*/ for /p/ in the onset of the second syllable. The correct response, then, is '틀리다/Incorrect'.

Perception Task 2: Nonword Identification

In the Nonword Identification task, learners hear a 1-2 syllable nonword and must select between two written options that differ by only one phoneme (Figure 4). Like the Nonword Reading task, nonwords take the form of (C/G)V, VC, or VCV. Here, learners' phonological knowledge is tapped via the decoding of the speech signal in the audio recording; this decoding is then compared to the phonological correspondences of the two written options. Naturally, orthographic knowledge is a potential confound in this task. Accordingly, printed response options are written to be maximally transparent, avoiding any sound-symbol mismatches. This task, like the Pronunciation Judgment task, is scored objectively, with 1 point awarded for correct answers and 0 for incorrect answers.

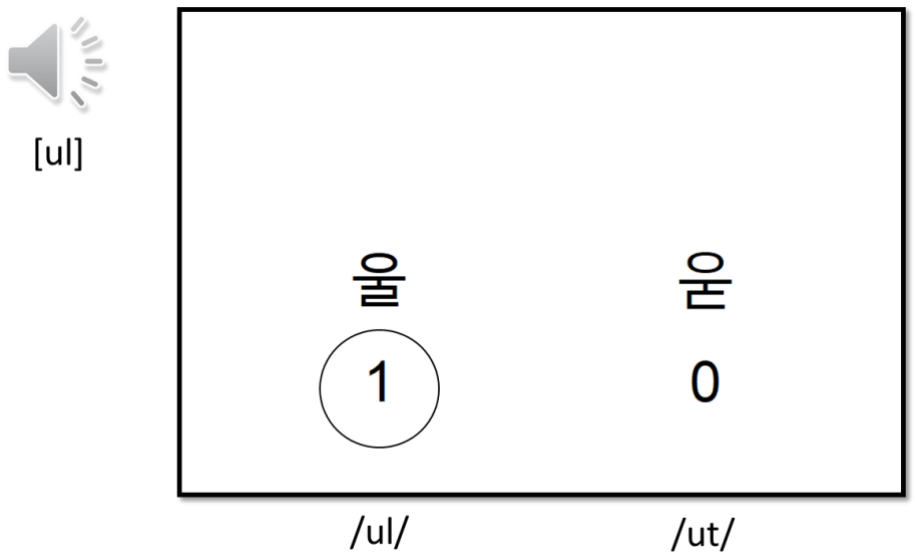

Figure 4. Schematic depicting the Nonword Identification task. In this example, learners hear the nonword [ul]. The correct response is 울 /ul/, on the left (cf. 웃, /ut/, on the right).

Diagnostic Score Reports

The role of KPD score reports is to provide detailed, instructionally-relatable, and digestible information on specific pronunciation strengths and weaknesses (Figure 5). The first page, shown on the left side of Figure 5, summarizes the learner's major pronunciation difficulties. The text at the top explains how to interpret the score report, and throughout the score report non-technical vocabulary is used whenever possible. No numeric scores are given on the first page. Instead, phonemes with low accuracy (e.g., < 75%) in production are flagged and listed, and learners are also informed of whether they had substantial difficulty perceiving those sounds. Supplementary information on difficult articulatory features and contexts (e.g., word-initial, word-final) is included and reported in a similar fashion. The intention is that learners (and teachers) focus on these sounds in their learning activity, rather than immediately overloading learners with a bevy of numeric scores and little guidance for what to do going forward. As Levis (2018) pointed out, the largest improvements to a learner's intelligibility are likely to come from a focus on a handful of the learner's most acute difficulties.

On the second page of the report (right side of Figure 5), numeric accuracy scores are provided for each phoneme in production and perception. This information may sate the curiosity of students, but more instructionally useful is the rightmost column, which contains real words (from the Picture Naming and Pronunciation Judgment tasks) that the learner responded to unclearly or incorrectly. These words could help with interpretation and/or retention of the results (e.g., a learner can know they have a difficulty with sound X in the common word Y) and may be useful in the initial generation of study materials.

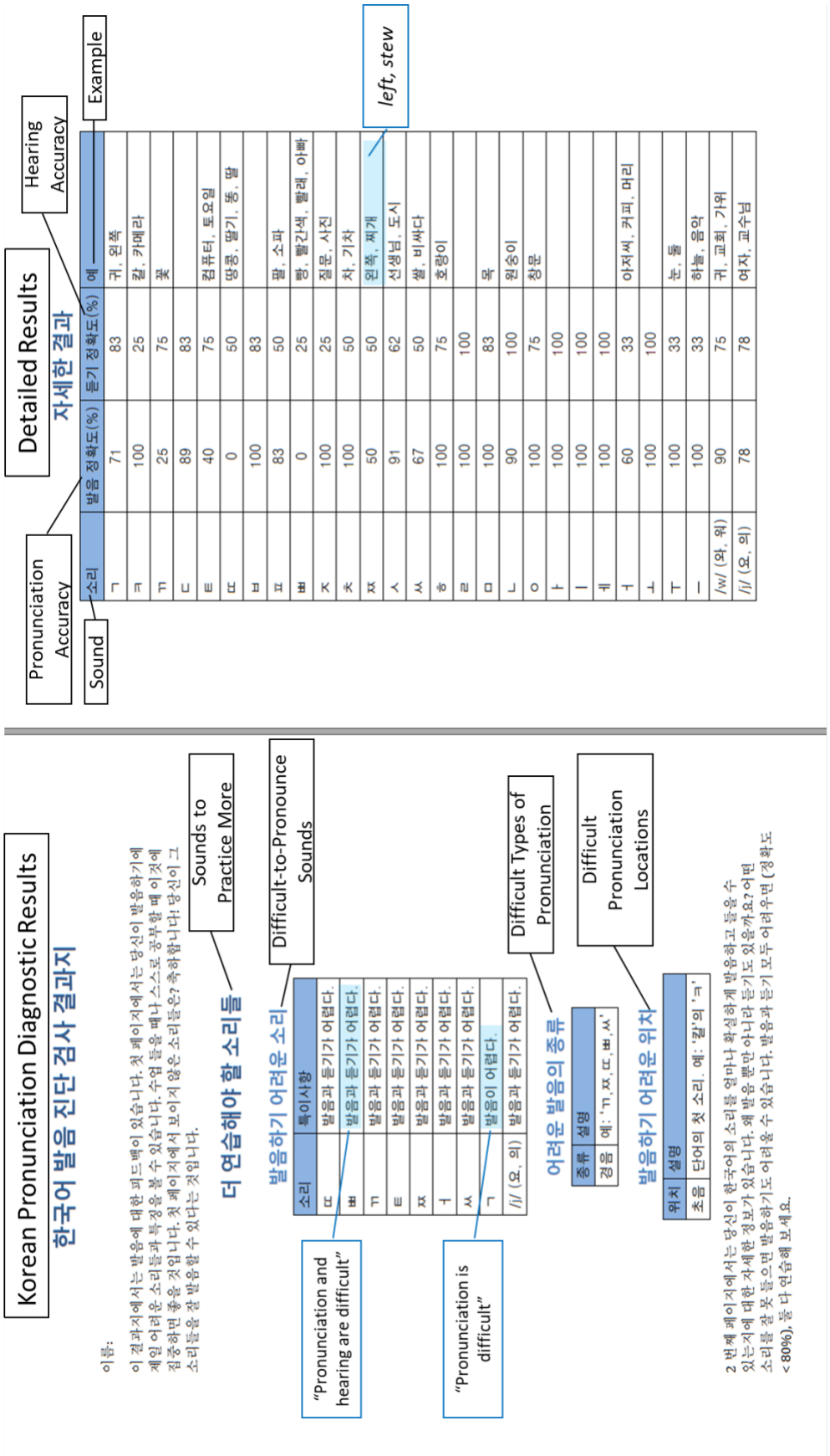

Figure 5. Diagram of a KPD score report. The first page of the score report is shown on the left, and the second page on the right.

INVESTIGATING THE VALIDITY OF A PRONUNCIATION DIAGNOSTIC

Designing and subsequently administering a pronunciation diagnostic is not the end of the story. Like all forms of assessment, validity is a chief concern: It would be of little use to administer a low-quality, invalid assessment.

Validity has come to mean more than just answering the question "does a test measure what we think it does?" Rather, the concept of validity includes the interpretations of test scores and resulting decisions and impacts of those decisions. Beyond measuring the targeted construct, other questions also have importance, like "Do the scores reflect performance in non-assessment situations?" and "Are decisions made based on scores leading to benefits for learners?" To help this expanded view of validity maintain coherence, assessment designers construct *interpretive arguments* (Chapelle et al., 2010) that trace the logical steps (inferences) and supporting evidence that start with features of test design and lead to learner responses, scores, and real-world decisions. The KPD's interpretive argument is shown in Figure 6. Interpretive arguments are customized based on purpose and context of the assessment (in this case, supporting individualized instruction for L2 Korean learners) and thus do not have an all-purpose template; however, those seeking to construct diagnostic tools for pronunciation in other languages could use the KPD's argument as a starting point.

An interpretive argument requires verification, and this is where validation research comes in. After designing and developing the KPD, I set out to investigate key assumptions that underlie the KPD's interpretive argument. These efforts included administering the KPD and collecting other information from 198 learners of Korean. It is beyond the scope of this paper to review the evidence which supports the validity of KPD score interpretations. However, the point to be made is that thought should be given to what happens after the use of an assessment tool or technique (e.g., a test, teacher observation, self-assessment), and that the interpretation and uses of test scores should be evaluated empirically.

Inference	Explanation
Usefulness & Impact	Appropriate application of score information by learners and teachers leads to improved intelligibility of segmental production and accurate segmental perception.
Utilization	Phoneme-level scores and diagnostic flags are interpretable and useful to learners and teachers for planning pronunciation learning activity and raising awareness of difficulties.
Extrapolation	The knowledge and abilities measured by the KPD are relevant to learner performances in general oral communication.
Explanation	Scores are reflective of learners' underlying phoneme knowledge.
Generalization	Observed scores estimate learners' abilities with stability and are similar across scorers.
Evaluation	Observations of phoneme production and perception are evaluated to yield scores that are instructionally relevant, indicative of strengths and weaknesses, and in line with intelligibility.
Operationalization	Observations of learners' Korean segmental production and perception reveal underlying strengths and weaknesses in phonological knowledge important to communication and relevant to development of intelligible pronunciation.

Figure 6. The interpretive argument of the KPD.

CONCLUSION

This article made the case for DLA in pronunciation teaching and learning and illustrated what form a DLA tool could take with an example targeting L2 Korean segmental pronunciation through the KPD. The KPD is distinct from other pronunciation assessments by its detailed coverage of Korean's phonemic inventory and inclusion of both perceptive and productive skills. As part of a larger project, the KPD was subjected to rigorous validation research. The KPD and my approach to investigating its validity may serve as a useful model, or at least point of reference, for others seeking to develop DLA tools for the pronunciation of other languages.

ACKNOWLEDGEMENTS

This project was supported by a Fulbright U.S. Student Junior Researcher fellowship and an Educational Testing Service TOEFL Small Grant for Dissertation Research.

ABOUT THE AUTHOR

Daniel R. Isbell (isbell.daniel@gmail.com) is an Assistant Professor in the Department of Second Language Study at the University of Hawai‘i at Mānoa. Dan’s research interests include language assessment, L2 pronunciation, and instructed SLA.

REFERENCES

Alderson, J. C. (2005). *Diagnosing foreign language proficiency: The interface between learning and assessment.* New York, NY: Continuum.

Alderson, J. C., Brunfaut, T., & Harding, L. (2015). Towards a theory of diagnosis in second and foreign language assessment: Insights from professional practice across diverse fields. *Applied Linguistics*, *36*(2), 236–260. https://doi.org/10.1093/applin/amt046

Amengual, M. (2016). The perception of language-specific phonetic categories does not guarantee accurate phonological representations in the lexicon of early bilinguals. *Applied Psycholinguistics*, *37*, 1221-1251. https://doi.org/10.1017/S0142716415000557

Brodeur, M. B., Dionne-Dostie, E., Montreuil, T., & Lepage, M. (2010). The bank of standardized stimuli (BOSS), a new set of 480 normative photos of objects to be used as visual stimuli in cognitive research. *PloS ONE, 5*(5), e10773.

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. https://doi.org/10.1111/j.1745-3992.2009.00165.x

Council of Europe. (2017). *Common European framework of reference for languages: Learning, teaching, assessment – Companion volume with new descriptors.* Retrieved from https://rm.coe.int/common-european-framework-of-reference-for-languages-learningteaching/168074a4e2

Crowther, D., Isaacs, T., Trofimovich, P., & Saito, K. (2015). Does a speaking task affect second language comprehensibility? *Modern Language Journal, 99*(1), 80-95. https://doi.org/10.1111/modl.12185

Derwing, T. M., & Munro, M. J. (2015). *Pronunciation fundamentals: Evidence-based perspectives for L2 teaching and research*. Philadelphia, PA: John Benjamins.

Duñabeitia, J.A., Crepaldi, D., Meyer, A.S., New, B., Pliatsikas, C., Smolka, E., & Brysbaert, M. (2017). MultiPic: A standardized set of 750 drawings with norms for six European languages. *Quarterly Journal of Experimental Psychology, 71(4), 808-816*. https://doi.org/10.1080/17470218.2017.1310261

Field, J. (2011). Cognitive validity. In L. Taylor (Ed.), *Examining speaking: Research and practice in assessing second language speaking* (pp. 65-111). Cambridge: Cambridge University Press.

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

Flege, J. E. (1991). Perception and production: The relevance of phonetic input to l2 phonological learning. In T. Hueber and C. Ferguson (Eds.), *Crosscurrents in second language acquisition and linguistic theories* (pp. 249-289). Amsterdam: John Benjamins.

Guion, S. G., & Pederson, E. (2007). Investigating the role of attention in phonetic learning. In O.-S. Bohn & M. J. Munro (Eds.), *Language Learning & Language Teaching* (Vol. 17, pp. 57–77). https://doi.org/10.1075/lllt.17.09gui

Harding, L., Alderson, J. C., & Brunfaut, T. (2015). Diagnostic assessment of reading and listening in a second or foreign language: Elaborating on diagnostic principles. *Language Testing*, *32*(3), 317–336. https://doi.org/10.1177/0265532214564505

Holliday, J. J., Turnbull, R., & Eychenne, J. (2017). K-SPAN: A lexical database of Korean surface forms and phonological neighborhood density statistics. *Behavior Research Methods*, *49*, 1939-1950. http://doi.org/10.3758/s13428-016-0836-8

Isaacs, T., & Trofimovich, P. (Eds.). (2017). *Second language pronunciation assessment: Interdisciplinary perspectives.* Bristol, UK: Multilingual Matters.

Kang, O., & Ginther, A. (Eds.). (2017). *Assessment in second language pronunciation.* New York: Routledge.

Kang, O., Thomson, R. I., & Moran, M. (2018). Which features of accent affect understanding? Exploring the intelligibility threshold of diverse accent varieties. *Applied Linguistics*. https://doi.org/10.1093/applin/amy053

Moyer, A. (2014). Exceptional outcomes in L2 phonology: The critical factors of learner engagement and self-regulation. *Applied Linguistics*, *35*(4), 418-440. https://doi.org/10.1093/applin/amu012

Lado, R. (1961). *Language testing: The construction and use of foreign language tests: A teacher's book.* London: Longmans, Green and Company.

Lee, A. H., & Lyster, R. (2017). Can corrective feedback on second language speech perception errors affect production accuracy? *Applied Psycholinguistics*, *38*, 371-393. https://doi.org/10.1017/S0142716416000254

Lee, S-H., Jang, S. B., & Seo, S. K. (2017). *A frequency dictionary of Korean.* New York: Routledge.

Levis, J. M. (2005). Changing contexts and shifting paradigms in pronunciation teaching. *TESOL Quarterly*, *39*(3), 369. https://doi.org/10.2307/3588485

Levis., J. M. (2018). *Intelligibility, oral communication, and the teaching of pronunciation.* Cambridge, UK: Cambridge University Press.

Levis, J. M. (2019). Cinderella no more: Leaving victimhood behind. *Speak Out! Journal of the IATEFL Pronunciation Special Interest Group, 60*, 7-14.

Levis, J., & Barriuso, T. A. (2012). Nonnative speakers' pronunciation errors in spoken and read English. In J. Levis & K. LeVelle (Eds.), *Proceedings of the 3rd Annual Pronunciation in Second Language Learning and Teaching Conference* (pp. 187-194). Ames, IA: Iowa State University.

Loewen, S., & Isbell, D. R. (2017). Pronunciation in face-to-face and oral synchronous computer mediated interaction. *Studies in Second Language Acquisition, 39*(2), 225-256. https://doi.org/10.1017/S0272263116000449

Oh, Y. M., Coupé, C., Marsico, E., & Pellegrino, F. (2015). Bridging phonological system and lexicon: Insights from a corpus study of functional load. *Journal of Phonetics, 53*, 153-176. https://doi.org/10.1016/j.wocn.2015.08.003

Saito, K. (2019). Individual differences in second language speech learning in classroom settings: Roles of awareness in the longitudinal development of Japanese learners' English /ɹ/ pronunciation. *Second Language Research, 35(2), 149-172.* https://doi.org/10.1177/0267658318768342

Sakai, M., & Moorman, C. (2018). Can perception training improve the production of second language phonemes? A meta-analytic review of 25 years of perception training research. *Applied Psycholinguistics, 39,* 187-224. https://doi.org/10.1017/S0142716417000418

Thomson, R. I. (2011). Computer Assisted Pronunciation Training: Targeting second language vowel perception improves pronunciation. *CALICO Journal*, *28,* 744-765.