Validity of the Proficiency in Oral English Communication Screening

Elizabeth S. Morton
Shelley B. Brundage
Adrienne B. Hancock
The George Washington University, Washington, DC

ABSTRACT: Purpose: This study investigated the construct, criterion, and social validity of the Proficiency in Oral English Communication Screening (POEC-S; Sikorski, 2005b), which is an assessment that is used with accented English speakers. Validity of this assessment has not previously been established despite its frequent use in clinical practice.

Method: Speech samples and scores from the POEC-S and the Test of English as a Foreign Language (TOEFL; Educational Testing Service, n.d.) were collected from 28 nonnative English speakers. Twenty unskilled listeners (undergraduate students) and 20 skilled listeners (speech-language pathologists) listened to the speech samples and rated each speaker on speech parameters of overall accent, articulation, intonation, naturalness, and estimated intelligibility using perceptual rating scales.

Results: The speakers’ POEC-S total and subtest scores correlated with the skilled listeners’ perceptual ratings of accent and, to a lesser extent, the speakers’ TOEFL scores; both results suggest the presence of construct validity. Criterion validity was demonstrated by significant correlations between the POEC-S scores and all components of our working definition of accent. Unskilled and skilled listener ratings also correlated with speakers’ POEC-S scores, verifying its social validity.

Conclusion: These results support the validity of the POEC-S to assess foreign accented speakers with high TOEFL scores.

KEY WORDS: validity, accent, spoken English, measurement, assessment

The foreign-born population in the United States increased by 57% from 1990 to 2000 (U.S. Bureau of the Census, 1990, 2000). In 2007, an estimated 12.6% of the U.S. population was foreign born, 84.4% of whom spoke a language other than English (U.S. Bureau of the Census, 2007). As accents are typically spoken by nonnative English speakers (American Speech-Language-Hearing Association [ASHA], 1998), these statistics evidence the increasing numbers of potential speakers of accented English in the United States. With the number of speakers with accents on the rise, the ability to define and assess accent becomes increasingly relevant to the field of speech-language pathology.

Nonnative speakers may wish to receive assessment and treatment to reduce the effects of their accent and to increase their communication effectiveness (ASHA, 2004). These types of elective services are within the scope of practice for speech-language pathologists (SLPs). SLPs have the professional knowledge and skills (e.g., the ability to distinguish a speech disorder from a speech difference) necessary to provide elective accent modification assessment and treatment to those who wish it (ASHA, 2004).

Assessing Accent

Evidence-based treatment is grounded in evidence-based assessment (ASHA, 2005). Schmidt and Sullivan (2003) surveyed university speech clinics in the United States to ascertain which tests SLPs used most frequently to evaluate and quantify foreign accents. These authors found that two tests were commonly used: Forty three percent of respondents used the Compton Phonological Assessment of Foreign Accent (CPAFA; Compton, 2002), and 37% used the Proficiency in Oral English Communication (POEC; Sikorski, 1991). The purpose of the CPAFA is to analyze the “accented sounds of American English for the foreign
The TOEFL is an English-language proficiency test and is administered by ETS (n.d.) for this purpose (ETS, 2009a, 2009b). However, more than 7,000 U.S. universities accept the Test of English as a foreign language (TOEFL) to assess English proficiency. The TOEFL takes 90–100 min to administer and score.

The purpose of the POEC is to assess “the extent and source of accent problems in the non-native speaker of American English” (Sikorski, 2007, p. 1). It consists of a comprehensive assessment and a screening version. The stated purpose of the screening version, the POEC-S (Sikorski, 2005b), is to “establish a baseline on the range and source of accent problems in the non-native speaker of American English” (Sikorski, 2005c, p. 1). The POEC tests define accent in a broad way that has not been examined by other assessment tools. Their unique blending of articulation, intonation, and auditory discrimination gives the POEC tests face validity, but their validity has not been tested empirically.

The POEC assessment takes approximately 1 hr to administer and at least 1 hr to score; the POEC-S takes 20–30 min to administer and can be scored during administration. The abbreviated administration time makes the POEC-S a practical screening tool. Before this study, these three tests, the CPAFA, POEC, and POEC-S, were similarly lacking in validity data. For the purposes of the current article, we chose to focus our efforts on the POEC-S due to its face validity and shorter administration time.

The POEC-S contains seven sections that measure a person’s auditory discrimination, intonation, and articulation (Sikorski, 2005b). Scores from all seven sections are added to produce a total possible score of 181 points. The first four sections of the POEC-S focus on auditory discrimination; the examinee is asked to discriminate between minimal pairs in words and sentences, write down words that he or she hears, and demonstrate knowledge of intonation by underlining the stressed word in sentences that are read aloud by the examiner. Section five assesses vowel accuracy by asking the examinee to read sentences aloud while the examiner marks whether the vowels were pronounced using Standard American English. In section six, the examinee reads questions and statements, and the examiner rates them on appropriate word stress and intonation contour. Finally, section seven tests articulation by asking the examinee to read phrases while the examiner rates parameters such as vowel length, voicing, final consonant production, and linking of words in conversational speech. An example of word linking occurs in the utterance Didn’t he. A native English speaker would not pronounce the /t/ or /h/ and would not pause between the two words. As a result, the utterance would be transcribed as /didint hi/. An optional conversational speech sample can be included in the assessment.

Foreign language proficiency testing is also available using tests not typically given by SLPs. For example, more than 7,000 U.S. universities accept the Test of English as a Foreign Language (TOEFL, Educational Testing Service [ETS], n.d.) for this purpose (ETS, 2009a, 2009b). However, the TOEFL is an English-language proficiency test and was not designed to specifically assess oral English skills.

**Accent**

Accent has been defined in many different ways and has been divided into various components in the linguistics and speech and hearing literature. Accent is defined by ASHA as “a phonetic trait from a person’s original language (L1) that is carried over [to] a second language (L2)” (1998, p. 28). According to this definition, foreign accent is composed strictly of phonetic differences.

A broader definition of accent can be found in Munro (1998), who defined accent as “nonpathological speech produced by second language (L2) learners that differs in partially systematic ways from the speech characteristic of native speakers of a given language” (1998, p. 139). Munro specifically mentions differences in phonemic production, intonation, and vocal quality as playing a role in the perception of accent. Other authors agree that articulation is one of many components of accent; however, accent also includes nonverbal skills; voice characteristics such as pitch, intensity, and resonance; and prosodic features including stress, intonation, coarticulation, and word linking (Gilbert, 1994; Pennington & Richards, 1986; Sikorski, 2005a, 2005b, 2005c; Wong, 1986). Intonation has been singled out as a key component of accent in the context of speech intelligibility (Gilbert, 1994; Sikorski, 2005a; Wong, 1986). Sikorski (2005a) emphasized that “non-standard intonation patterns” appear to negatively affect intelligibility more so than the other aspects of accent discussed earlier.

Intelligibility has been emphasized as an essential parameter involved in communication competence and accent training (ASHA, 2004; Morley, 1994).

Although not part of accent’s definition, “intelligibility” of speech is a construct that appears often in discussions of accent. Acoustic speech signals can be intelligible and/or comprehensible. Intelligibility is defined as “a decision by the listener that specifies how well the message was understood” (Speaks, Parker, Harris, & Kuhl, 1972, p. 592). A similar definition was used in a study of intelligibility of synthesized speech (Hustad, Kent, & Beukelman, 1998). According to Yorkston, Strand, and Kennedy (1996), intelligibility involves the speech signal and any of the speaker’s compensatory strategies. On the other hand, comprehensibility involves these two things and syntactical and semantic context, situational cues, and gestures (Yorkston et al., 1996). Although there is no one “universal” method for measuring intelligibility (Munro & Derwing, 1999), intelligibility can be measured reliably by having listeners use interval scales, direct magnitude estimation, orthographic transcription, or percentage ratings (Derwing & Munro, 1997; Southwood & Flege, 1999; Speaks et al., 1972). Whereas orthographic transcription provides an exact measure of intelligibility, asking listeners to estimate the percentage of a given speech passage that they understood improves measurement efficiency without sacrificing reliability (Darley, Aronson, & Brown, 1969; Speaks et al., 1972; Yorkston & Beukelman, 1978).

Measures of intelligibility may be influenced by some of the same variables that affect measures of accent, including contextual factors (e.g., single words vs. connected discourse), the signal-to-noise ratio in which speech samples...
are heard (e.g., quiet vs. noise), and the degree to which a given speaker’s speech differs from an accepted “norm” (Munro, 1998; Speaks et al., 1972; Yorkston & Beukelman, 1978). Because accent and intelligibility are by definition perceptual events that are influenced by similar external factors, it seems reasonable that they might be correlated in some way. Indeed, vanWijngaarden, Steeneeken, and Hout gast (2002) found significant positive correlations between ratings of accent and intelligibility scores for their heterogeneous sample of nonnative speakers of Dutch. These authors interpreted their findings to indicate that “the overall effect on speech intelligibility is proportional to the degree of foreign accent” (vanWijngaarden et al., 2002, p. 3.012). However, in a study using groups of native Mandarin and native English speakers, Munro and Derwing (1999) found that intelligibility was not strongly correlated with accent ratings. In this study, an intelligibility score was obtained by calculating the number of words that were correctly transcribed by listeners after hearing each speech sample. Listeners were also asked to rate each speaker’s strength of foreign accent using a 9-point Likert-type scale. Ratings of accent and intelligibility were significantly negatively correlated for 28% of the listeners. These studies had differences related to the samples (i.e., languages studied) and the way intelligibility was measured (i.e., transcription vs. perception). These differences may account for the different results obtained. Taken together, these studies suggest that disagreement exists in the literature about the relationship between accent and intelligibility.

A related, but distinct, measure used in evaluating speech output is speech naturalness (Hustad et al., 1998). Naturalness has been used as an outcome measure in studies of speech disorders such as stuttering and dysarthria (Hearne, Packman, Onslow, & O’Brien, 2008; Martin, Haroldson, & Triden, 1984; Tasko, McClean, & Runyan, 2007). Griffen (1991) suggested that speech naturalness and accent components are related when he said that “the goal of instruction (1991) suggested that speech naturalness and accent components are related when he said that “the goal of instruction in pronunciation is that the student (or patient) should learn to speak the language as naturally as possible” (p. 182). Relationships between naturalness and accent were also addressed by Mackey, Finn, and Ingham (1997), who found that naive listeners who were General American English speakers rated accented speakers as sounding significantly less natural than Standard American English speakers. In Mackey et al.’s study, naturalness was specifically left undefined, with only highly natural and unnatural as anchors on a 9-point rating scale.

Although some researchers (Schiavetti & Metz, 1997), have called for a standard definition of naturalness to be adopted for use in research, naturalness is a construct that has been reliably rated in previous studies without providing definitions of it (Eadie & Doyle, 2002; Martin et al., 1984; Mackey et al., 1997). Furthermore, some researchers have been able to identify components of naturalness, including voice onset time, phonation duration, sentence duration, rate, rhythm, intonation, stress patterns, and fluency (Finn & Ingham, 1994; Gow & Ingham, 1992; Martin & Haroldson, 1992; Martin et al., 1984; Metz, Schiavetti, & Sacco, 1990; Onslow, Hays, Hutchins, & Newman, 1992; Yorkston, Beukelman, & Bell, 1988). Therefore, although a standard definition of naturalness has not yet been adopted, there is evidence to indicate that several speech features play a role in naturalness. It is feasible that naturalness and accent are related, as prosodic features have been linked to defining both.

To summarize the definitions and components of accent discussed so far, accent can include differences in articulation, prosody (i.e., intonation, word linking, stress, rhythm), vocal quality, fluency, grammar, and nonverbal skills. The literature is conflicting regarding the relationship between accent and intelligibility. In contrast, ratings of degree of accent are greatly consistently inversely correlated with ratings of speech naturalness. Thus, SLPs endeavoring to assist persons in reducing their level of accentedness might reasonably measure changes in articulation and intonation over time, for example, and quantify the positive changes in intelligibility and naturalness that accompany these articulatory and intonation changes.

Validity

McCauley and Swisher (1984) suggested that there are four types of validity to consider when evaluating assessments: content validity, face validity, construct validity, and criterion validity. Content validity is determined when an expert on the topic deems the test appropriate based on examination of the test items and their apparent relevance to the topic. It is important to note that this type of opinion cannot suffice as evidence for how well or to what extent the instrument’s content measures accent. Face validity is similar to content validity except that it involves evaluation of the test by an untrained individual, so it is considered to be a lower standard of validity. Both content and face validity are considered to be qualitative assessments and are not quantified or analyzed statistically (Schiavetti & Metz, 2006). Frequent use of the CPAFA and POEC by SLPs indicates acceptable content and face validity of these measures. However, part of content validity is that the test items represent the entire range of possible items the test should cover to measure the construct. The differences between the two most commonly used tests, CPAFA and POEC, suggest at least two views of what content is important in quantifying accent.

Construct validity examines the association between a test score and the prediction of a theoretical characteristic, such as intelligence or accent (Cozby, 2007; Schiavetti & Metz, 2006). Currently, an SLP’s perceptual judgment of accent is the gold standard for determining the degree of a person’s overall accent, and the TOEFL is a broader, more objective measure of essentially the same construct with some limitations. Therefore, correlations between a POEC-S score and these two measures (i.e., SLP’s rating of “overall accent” and TOEFL score) would indicate construct validity of the POEC-S.

Finally, criterion validity involves comparing an individual’s score on the assessment being studied to another established method of assessing the same parameter(s) (McCauley & Swisher, 1984). The established method or parameters are called the criterion, and it can be a
standardized assessment or an expert’s assessment. For example, if a new child language assessment was being studied for criterion validity, a researcher could compare a client’s scores on the new assessment to his or her scores on an already established method of assessing child language, which is considered to be an indirect measurement. Alternatively, the researcher could compare scores on the new assessment to an SLP’s judgment of the child’s language, which would be a direct measurement.

In order to assess the criterion validity of the POEC-S, the current study will use a direct measure of assessing validity. Because a valid assessment for measuring oral English proficiency has not yet been established, criterion validity cannot be assessed indirectly via comparison to an established assessment instrument. The TOEFL scores available for this study measured listening, reading, and writing, so they cannot be used to establish the POEC-S’s criterion validity indirectly. It will, therefore, be assessed directly by comparing speakers’ POEC-S scores to SLPs’ ratings of components of accent. The literature review indicates that articulation, intonation, naturalness, and intelligibility are all criteria used in measuring accent.

As communication specialists, SLPs have unique skills that enable them to provide services in assessment and intervention for accent modification in nonnative English speakers. As part of their clinical education, SLPs are trained in perceptual measurement for a variety of communication differences and disorders. These perceptual skills are important when evaluating tests of accent because, as Southwood and Feige (1999) noted, “identifying speaker differences is a perceptual event, and the listeners’ reactions validate any measurements used to determine such differences” (p. 336). Additionally, SLPs have extensive training in articulation and phonological assessment and intervention, voice and intonation, cultural sensitivity, and identification of normal versus disordered speech. This combination of skills indicates that SLPs are skilled in assessing and training accent modification clients (Sikorski, 2005d). For these reasons, we consider SLPs to be “skilled” listeners. Therefore, SLPs’ ratings of overall accent can be used to determine construct validity, and SLPs’ ratings of articulation, intonation, naturalness, and estimated intelligibility can be used to document criterion validity.

A final, important type of validity was discussed by Wolf (1978). He argued that social validity has been overlooked in the past but should be included more frequently in research. He defined social validity as the degree of significance that a program has to its consumers. Therefore, a socially valid assessment would have a great deal of significance to its consumers, or would perhaps correlate with the opinions of the consumers. In this study, undergraduate students (hereafter, “unskilled listeners”) could be considered consumers as they regularly encounter and interact with individuals with foreign accents as part of their university education. Therefore, if the POEC-S has social validity, the scores obtained on the POEC-S would correlate with the unskilled listeners’ ratings of the oral English skills of the individuals being tested. It is valuable to collect several measures related to social acceptance of speech because it is possible to improve one at the expense of another (e.g., become more intelligible but be less natural). Fawcett (1991) stated that rating scales are appropriate for assessing social validity; therefore, unskilled listener ratings of speech samples were used in the current study as a measure of social validity.

## Purpose and Research Questions

The purpose of the present study was to determine whether the POEC-S is a valid measure of accent. This study had several research questions. The first two questions refer to construct validity:

- Do skilled listeners’ ratings of overall accent correlate with the speakers’ POEC-S scores?
- Do scores obtained on the TOEFL, an English-language proficiency test, correlate with scores obtained on the POEC-S?

The third question addresses criterion validity:

- Do skilled listeners’ ratings of articulation, intonation, naturalness, and estimated intelligibility correlate with the speakers’ POEC-S scores and subtest scores?

The fourth question addresses social validity:

- Do perceptual ratings made by unskilled listeners correlate with the speakers’ POEC-S scores?

The authors hypothesized that the skilled listeners’ ratings of degree of accent and their ratings of articulation, intonation, naturalness, and estimated intelligibility would correlate with the speakers’ POEC-S scores, as the POEC-S is composed of sections relating to these areas and does appear to measure degree of accent. It was hypothesized that the English-language proficiency test scores (TOEFL scores) would be less strongly correlated with the POEC-S scores, as TOEFL scores are composite scores that represent a broad range of English-language skills, including listening, speaking, reading, and writing. The authors also expected the POEC-S to have social validity, as demonstrated by a relationship between unskilled listeners’ perceptions and speakers’ POEC-S scores.

## METHOD

### Participants: Listeners

There were two groups of listeners: unskilled listeners and skilled listeners. The unskilled listeners group consisted of 20 undergraduate students ranging in age from 18 to 27 years, with a mean age of 20 years. There were 19 females and one male who were recruited through flyers and classroom recruitment.

The skilled listeners group consisted of 20 SLPs with ASHA’s Certificate of Clinical Competence (CCC). These listeners ranged in age from 25 to 53 years, with a mean age of 36 years, and included 19 females and one male. The number of self-reported years of experience after having completed a clinical fellowship year (CFY) for the SLPs ranged from 0 to 25. The skilled listeners were recruited through e-mails and through professional
contacts from the Speech and Hearing Sciences Department at George Washington University (GWU).

Inclusion criteria required each participant to have normal hearing, be between the ages of 18 and 70 years, and be a native English speaker. All entrance criteria were determined by self-report, with the exception of hearing thresholds. All participants were required to pass a hearing screening at 500 Hz, 1000 Hz, 2000 Hz, and 4000 Hz presented at 20 dB bilaterally. No participant reported having any speech or language disorders. All of the listeners were native speakers of English and were not paid for participating in the study. The study was reviewed and approved by GWU’s Institutional Review Board.

Participants: Speakers

The POEC-S and a spontaneous speech sample were administered to 28 newly enrolled international graduate students who had been awarded teaching assistantships. There were 15 males and 13 females, and all were nonnative English speakers. The average age of the speakers was 26 years (SD = 3.8, range = 21–37 years). The average self-reported number of years speaking English was 14 (SD = 6.8, range = 1–28 years). The native languages of these speakers included Arabic, Chinese, Farsi, French, Greek, Hindi, Igbo, Japanese, Portuguese, Romanian, Serbian, Shona, Slovenian, Spanish, and Turkish. This diversity of accents is consistent with those studied in prior linguistic studies (Anderson-Hsieh, Johnson, & Koehler, 1992; Mackey et al., 1997; vanWijngaarden et al., 2002).

Materials and Speech Stimuli Development

All speakers who participated in stimuli development were referred to a university speech and hearing clinic by the university fellowship office for an oral English proficiency screening before commencing their teaching duties. Informed consent was obtained from each speaker before any research procedures began. Thirty speakers signed consent forms, but two female participants failed their hearing screenings, which resulted in 28 speakers for the study. Speakers were unpaid, and participation in the study had no bearing on their assistantships/fellowships.

Although the only entrance criteria for speakers in this study were that they be nonnative English speakers, be graduate students at GWU, and have normal hearing, it is important to note that students who received teaching assistantships had already been screened through the university admissions process. In order to gain admission into a graduate program, the students were required to have strong academic records. Additionally, in order to be considered for a teaching assistantship, they were required to score at least 100 on the Internet-based TOEFL, 250 on the computer-based TOEFL, 600 on the paper-based TOEFL, or 7.0 on another less commonly used English language assessment entitled the International English Language Testing System (n.d.). These scores correlate to competence in the English language across comprehension and expression for social, professional, and academic purposes (Tannenbaum & Wylie, 2005). Each speaker then signed a release form that allowed the university to release his or her TOEFL score report to the researchers. Six speakers either did not have or did not release their TOEFL scores. However, these speakers were kept in the study despite not having this information.

Each speaker filled out a demographic questionnaire, underwent a hearing screening, and was administered the POEC-S. These assessments took approximately 1 hr to complete. When the POEC-S was administered to each speaker, sections 1–4 were prerecorded and administered via an audio CD. Sections 5–7 were administered by an SLP in her CFY.

Additionally, speakers were instructed to provide a 2- to 3-min speech sample about a book, movie, or television program. The task was somewhat controlled and was designed to result in common vocabulary and terminology while still leaving room for individual differences and vocabulary variation. If the topic had been the speaker’s field of study, raters may have altered their ratings based on jargon or unfamiliar terminology. These speech samples were all recorded using a CD recorder. The same CD recorder and microphone were used for all recordings. The Shure Beta 58A microphone was positioned the same distance from each speaker during all recordings of speech samples.

Each speaker provided a POEC-S score consisting of individual scores on all seven subsections. Additionally, 22 of the 28 speakers provided TOEFL scores. These materials were used in the rating and data analysis portions of the current study.

Preparing Speech Samples for Playback to Listeners

A 30-s portion of each speech sample recording was chosen for use in perceptual analyses by unskilled and skilled listeners. The samples were chosen by skipping the first 20 s of the recorded sample and using the next 30 s. This rule was used as long as examiner interjections or pauses longer than 5 s did not occur during the 30-s sample. When interjections or long pauses occurred during the 30-s sample, the first continuous 30-s sample within the recording was used for the rating portion of the study.

Seven of the 35 speech samples were repeated samples for intrarater reliability purposes. These samples were chosen to represent a range of accent degree based on POEC-S scores. Two of the repeated samples were from individuals who had high POEC-S scores, three were from individuals with moderate scores, and two were from individuals with low scores compared to the group as a whole. Therefore, the seven samples that were repeated were considered to be representative of the whole sample.

Rating Procedure

Unskilled and skilled listeners underwent the same procedure. First, each listener was given an information sheet that contained information about the study. The sheet was reviewed with the examiner, and each listener was given
the opportunity to ask questions. Next, each listener filled out a demographic questionnaire before undergoing a hearing screening at 500 Hz, 1000 Hz, 2000 Hz, and 4000 Hz presented bilaterally at 20 dB.

Next, instructions were read, and each listener was presented with two practice samples at a comfortable listening level. The practice samples were the same length as the test samples, and listeners were asked to make the same ratings for the practice samples as for the test samples. Each sample was 30 s long. Each listener was instructed to listen without writing; once the sample ended, each listener rated the speech across five parameters: overall accent, articulation, intonation, naturalness, and estimated intelligibility. The accent, articulation, intonation, and naturalness ratings were made on a 9-point Likert scale. For accent, 1 represented no accent and 9 represented profound accent. For articulation and intonation, 1 represented poor and 9 represented excellent. For naturalness, 1 was defined as very unnatural and 9 was defined as very natural. The estimated intelligibility ratings were based on percentages, where listeners circled the percentage range in 10% increments that represented how much of the sample they were able to understand. Both unskilled listeners and skilled listeners used the same rating form.

Each participant rated all 28 test samples across all five parameters and additionally rated seven of those samples for reliability purposes (35 samples total). The listeners were not told that samples would be repeated, and two identical samples were never played in proximity to each other. Presentation was counterbalanced for listening sequences within each listener group (i.e., unskilled and skilled); that is, half of the listeners in each group heard samples 1–35 followed by samples 18–35, and the other half of the listeners heard the seven samples that were repeated. For each group of unskilled and skilled listeners by comparing the ratings of the seven samples that were repeated. For each group of listeners across each parameter rated, the difference was calculated between the first and second ratings for each speech sample using the following equation: Time 2 – Time 1 = difference. For example, if a listener assigned a speech sample a 5 for accent on the first rating and then gave a 7 for accent the second time she heard that sample, the difference of 2 was calculated (7 – 5 = 2). If the difference was ≤ 1, it was considered to be in agreement. For example, if a listener gave a speech sample a 5 for accent on the first rating, then a rating of 4, 5, or 6 given on the second rating was considered to be in agreement. Inspection of the data revealed the presence of possible outliers, so the mean was calculated for the difference values for each parameter and each group. Any difference value that was > 2 SDs from the mean difference in either direction was considered to be an outlier. Outliers that met this definition were removed from the sample. In total, 94 out of 1,400 ratings (6.7%) were removed as outliers. There did not appear to be a pattern to the outliers in that they were not related to particular speakers or speech samples. The percentages of intrarater reliability after outliers were included in the study. The POEC-S consists of an objective auditory discrimination portion worth 31 points and a perceptually measured verbal portion worth 150 points. The 31 points given for auditory discrimination skills are scored objectively based on items that the examinee marks correctly or incorrectly, and the 150 points given for verbal skills are scored based on the examiner’s perception of the examinee’s speech. Point-to-point percentage agreement was calculated on both the verbal portion (150 points) and the total test (181 points). The point-to-point percentage agreement for the total POEC-S scores between the two SLPs was 93%. The point-to-point percentage agreement for the verbal POEC-S scores between the two SLPs was 91%.

Although the two SLPs scored the actual test items, the first author computed the scores and entered them into Microsoft Excel. A graduate student teaching assistant also computed and entered a subset of 10 out of the 28 (35%) speakers’ complete scores into Excel. When compared point-to-point, the agreement across these calculations and entries was 99.6%. The researcher referred back to the original score forms in order to resolve all discrepancies. Other data entered into Excel included the information collected from the demographic questionnaires, all unskilled and skilled listeners’ ratings, and the speakers’ TOEFL scores. TOEFL scores were converted and entered by both the researcher and the graduate teaching assistant. All of these data were entered by a member of the research team as well as by a teaching assistant.

The demographic information data entry agreement was 99.8%. The agreement on entering the unskilled listeners’ ratings and skilled listeners’ ratings was 99.7%. The TOEFL score entry and conversion agreement was 100%. For all items that were not in agreement on data entry reliability, the original demographic and rating forms were consulted in order to determine the true values for these items. In this way, any discrepancies were resolved.

Finally, intrarater reliability was calculated for the unskilled and skilled listeners by comparing the ratings of the seven samples that were repeated. For each group of listeners across each parameter rated, the difference was calculated between the first and second ratings for each speech sample using the following equation: Time 2 – Time 1 = difference. For example, if a listener assigned a speech sample a 5 for accent on the first rating and then gave a 7 for accent the second time she heard that sample, the difference of 2 was calculated (7 – 5 = 2). If the difference was ≤ 1, it was considered to be in agreement. For example, if a listener gave a speech sample a 5 for accent on the first rating, then a rating of 4, 5, or 6 given on the second rating was considered to be in agreement. Inspection of the data revealed the presence of possible outliers, so the mean was calculated for the difference values for each parameter and each group. Any difference value that was > 2 SDs from the mean difference in either direction was considered to be an outlier. Outliers that met this definition were removed from the sample. In total, 94 out of 1,400 ratings (6.7%) were removed as outliers. There did not appear to be a pattern to the outliers in that they were not related to particular speakers or speech samples. The percentages of intrarater reliability after outliers were included in the study. The POEC-S consists of an objective auditory discrimination portion worth 31 points and a perceptually measured verbal portion worth 150 points. The 31 points given for auditory discrimination skills are scored objectively based on items that the examinee marks correctly or incorrectly, and the 150 points given for verbal skills are scored based on the examiner’s perception of the examinee’s speech. Point-to-point percentage agreement was calculated on both the verbal portion (150 points) and the total test (181 points). The point-to-point percentage agreement for the total POEC-S scores between the two SLPs was 93%. The point-to-point percentage agreement for the verbal POEC-S scores between the two SLPs was 91%.

Although the two SLPs scored the actual test items, the first author computed the scores and entered them into Microsoft Excel. A graduate student teaching assistant also computed and entered a subset of 10 out of the 28 (35%) speakers’ complete scores into Excel. When compared point-to-point, the agreement across these calculations and entries was 99.6%. The researcher referred back to the original score forms in order to resolve all discrepancies. Other data entered into Excel included the information collected from the demographic questionnaires, all unskilled and skilled listeners’ ratings, and the speakers’ TOEFL scores. TOEFL scores were converted and entered by both the researcher and the graduate teaching assistant. All of these data were entered by a member of the research team as well as by a teaching assistant.

The demographic information data entry agreement was 99.8%. The agreement on entering the unskilled listeners’ ratings and skilled listeners’ ratings was 99.7%. The TOEFL score entry and conversion agreement was 100%. For all items that were not in agreement on data entry reliability, the original demographic and rating forms were consulted in order to determine the true values for these items. In this way, any discrepancies were resolved.

Finally, intrarater reliability was calculated for the unskilled and skilled listeners by comparing the ratings of the seven samples that were repeated. For each group of listeners across each parameter rated, the difference was calculated between the first and second ratings for each speech sample using the following equation: Time 2 – Time 1 = difference. For example, if a listener assigned a speech sample a 5 for accent on the first rating and then gave a 7 for accent the second time she heard that sample, the difference of 2 was calculated (7 – 5 = 2). If the difference was ≤ 1, it was considered to be in agreement. For example, if a listener gave a speech sample a 5 for accent on the first rating, then a rating of 4, 5, or 6 given on the second rating was considered to be in agreement. Inspection of the data revealed the presence of possible outliers, so the mean was calculated for the difference values for each parameter and each group. Any difference value that was > 2 SDs from the mean difference in either direction was considered to be an outlier. Outliers that met this definition were removed from the sample. In total, 94 out of 1,400 ratings (6.7%) were removed as outliers. There did not appear to be a pattern to the outliers in that they were not related to particular speakers or speech samples. The percentages of intrarater reliability after outliers were
removed are displayed in Table 1. As the reliability sample was considered to be representative of the sample as a whole, it can be assumed that these percentages of intrarater reliability would carry over to the rest of the sample. A criterion for intrarater reliability was set at 80%. The unskilled listeners’ ratings of accent and naturalness did not meet this criterion. Therefore, these ratings were removed from the data and were not used in further calculations.

RESULTS

Although counterbalancing was used to negate any order effects, an analysis was performed to determine if the order of presentation of the speech samples affected the ratings. Listeners in each group (i.e., unskilled and skilled) were equally divided into two groups: A and B. The second half of Group A’s listening samples was heard as the first half by Group B. Independent-samples t tests revealed that unskilled listener Groups A and B were not significantly different for ratings of articulation or intonation, and skilled listener Groups A and B were not significantly different for ratings of articulation, intonation, naturalness, or estimated intelligibility. However, independent-samples t tests revealed significant differences for unskilled listeners’ ratings of estimated intelligibility (p = 0.01) between Groups A and B and skilled listeners’ ratings of accent (p = 0.04). It is unclear why these two parameters were affected by order of presentation. It does not appear that the speech samples themselves would have caused these differences, as other parameters rated were not affected by order. Although it is interesting to note that the order of the sample may play a role in certain accent-related perceptual measures, the counterbalancing employed in the current study should have sufficiently prevented these effects from influencing the results.

Additionally, it should be noted that the POEC-S results in an auditory discrimination score, which is worth a maximum of 31 points; a verbal score, which is worth a maximum of 150 points; and a total score, which is the sum of the auditory discrimination and verbal scores. Both the POEC-S verbal score and the total POEC-S score were used in calculations for the current study. The means and standard deviations of all variables analyzed in this study are listed in Table 2. SPSS 15.0 (2006) was used for all statistical calculations. Results are summarized in Table 3.

Do Skilled Listeners’ Ratings of Overall Accent Correlate With the Speakers’ POEC-S Scores?

Speakers’ total POEC-S scores and verbal POEC-S scores were compared to skilled listeners’ accent ratings using a Pearson product–moment correlation. As shown in Table 3 and Figure 1, skilled listeners’ accent ratings significantly correlated with speakers’ POEC-S total scores, r = −0.83, p ≤ 0.000, and with their POEC-S verbal scores, r = −0.81, p ≤ 0.000, at the 0.01 alpha level. Although this correlation demonstrates an inverse relationship between accent ratings and POEC-S scores, higher scores on the POEC-S are indicative of a relatively mild accent, whereas higher accent ratings on the perceptual rating scale represent a relatively severe accent. Therefore, in this case, an inverse relationship indicates that a strong accent according to POEC-S score is associated with a profound accent rating by skilled listeners.

Do Scores Obtained on the TOEFL Correlate With Scores Obtained on the POEC-S?

Speakers’ TOEFL scores were compared to their POEC-S scores using a Pearson product–moment correlation. As shown in Figure 2, speakers’ TOEFL scores correlated significantly with their POEC-S total scores, r = 0.78, p ≤ 0.000, and their POEC-S verbal scores, r = 0.74, p ≤ 0.000. High TOEFL scores represent stronger language skills, and low TOEFL scores represent weaker language skills. Therefore, a strong accent according to POEC-S score is associated with weak language skills measured by the TOEFL.

Table 2. Means and standard deviations of scores from the Proficiency in Oral English Communication Screening (POEC-S; Sikorski, 2005b; n = 28), and the Test of English as a Foreign Language (TOEFL; Educational Testing Service, n.d.; n = 22), and perceptual ratings of accent, articulation, intonation, naturalness, and intelligibility (unskilled listeners n = 20; skilled listeners n = 20)

Table 1. Intrarater reliability for ratings of accent, articulation, intonation, naturalness, and intelligibility for unskilled and skilled listeners.

<table>
<thead>
<tr>
<th>Group</th>
<th>Parameter</th>
<th>Percentage intrarater reliability</th>
<th>Met 80% reliability?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unskilled</td>
<td>Accent</td>
<td>74</td>
<td>No</td>
</tr>
<tr>
<td>Unskilled</td>
<td>Articulation</td>
<td>85</td>
<td>Yes</td>
</tr>
<tr>
<td>Unskilled</td>
<td>Intonation</td>
<td>85</td>
<td>Yes</td>
</tr>
<tr>
<td>Unskilled</td>
<td>Naturalness</td>
<td>78</td>
<td>No</td>
</tr>
<tr>
<td>Unskilled</td>
<td>Intelligibility</td>
<td>95</td>
<td>Yes</td>
</tr>
<tr>
<td>Skilled</td>
<td>Accent</td>
<td>82</td>
<td>Yes</td>
</tr>
<tr>
<td>Skilled</td>
<td>Articulation</td>
<td>88</td>
<td>Yes</td>
</tr>
<tr>
<td>Skilled</td>
<td>Intonation</td>
<td>87</td>
<td>Yes</td>
</tr>
<tr>
<td>Skilled</td>
<td>Naturalness</td>
<td>88</td>
<td>Yes</td>
</tr>
<tr>
<td>Skilled</td>
<td>Intelligibility</td>
<td>95</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Note. Unskilled listeners were undergraduate students, and skilled listeners were certified SLPs.
Do Skilled Listeners’ Ratings of Articulation, Intonation, Naturalness, and Estimated Intelligibility Correlate With the Speakers’ POEC-S Scores and Subtest Scores?

A Pearson product–moment correlation was conducted in order to answer this question. The skilled listeners’ ratings of articulation, intonation, naturalness, and estimated intelligibility were each found to correlate significantly with both the speakers’ POEC-S total scores and their POEC-S verbal scores (see Table 3). The POEC-S has subtests for articulation and intonation, but not for estimated intelligibility or naturalness. The POEC-S articulation subtest score for articulation was derived by adding the scores for POEC-S section 5, which measures vowel production, and section 7, which measures articulation variations. These POEC-S articulation subtest scores were found to significantly correlate with skilled listeners’ articulation ratings, \( r = 0.82, p \leq 0.000 \), and overall accent ratings, \( r = -0.79, p \leq 0.000 \). The POEC-S intonation subtest correlated with skilled listeners’ intonation ratings, \( r = 0.46, p = 0.015 \), but not with their overall accent ratings, \( r = 0.37, p = 0.59 \).

Do Perceptual Ratings Made by Unskilled Listeners Correlate With the Speakers’ POEC-S Scores?

For social validity, the authors were specifically concerned with perceptions of speech accent and estimated intelligibility. However, unskilled listeners’ ratings of accent and naturalness proved unreliable, and therefore could not be used in this analysis. Unskilled listeners’ ratings of estimated intelligibility were reliable, and therefore these ratings were used to evaluate social validity.

Unskilled listeners’ estimated intelligibility ratings significantly correlated with the speakers’ POEC-S total scores, \( r = 0.82, p \leq 0.000 \) (see Figure 3), and with their POEC-S verbal scores, \( r = 0.78, p \leq 0.000 \), suggesting that the POEC-S has social validity.

Summary of Results

The current study sought to determine whether the POEC-S is a valid measure of accent. POEC-S scores correlated significantly with both measures used here to support construct validity (i.e., TOEFL scores and skilled listeners’ “overall accent” ratings) and the validating criterion derived from the literature of articulation, intonation, naturalness, and estimated intelligibility. The latter significant correlations support criterion validity of the POEC-S. Furthermore, estimated intelligibility ratings by unskilled listeners correlated with POEC-S scores, supporting social validity of the POEC-S.

DISCUSSION

Clinical Implications of POEC-S Validity

The current study supports construct, criterion, and social validity for the POEC-S. Results of this study suggest that the POEC-S is a valid measurement of accent. Speakers’ POEC-S total scores and POEC-S verbal scores correlated strongly with skilled listeners’ perceptual ratings of accent,
articulation, intonation, naturalness, and estimated intelligibility. These strong relationships between POEC-S scores and perceptual measures are important because perceptual measurements are an essential and primary tool in the field of speech-language pathology (Darley, 1984), in part due to their convenience and economy (Kent, 1996). Furthermore, in many instances in speech-language pathology, perceptual measurement is the only available method of measurement (Kent, 1996). Before the current study, validity had not been established for the POEC-S, which is a test that is used by SLPs to evaluate accent. Our findings suggest that the POEC-S has construct and criterion validity for the population measured in this study. Social validity was also established for the POEC-S because unskilled listeners’ estimated intelligibility ratings correlated with the speakers’ POEC-S scores.

This study also yielded interesting findings regarding specific subtests of the POEC-S. The subtest scores for intonation and articulation were shown to have criterion validity, as scores on these subtests correlated with skilled listeners’ perceptual ratings of intonation and articulation. However, the correlation between the intonation subtest and the intonation ratings was markedly weaker than the other intonation correlations calculated in the current study. Interestingly, the POEC-S total scores and POEC-S verbal scores were more predictive than the intonation subtest scores of listeners’ intonation ratings. Additionally, the intonation subtest scores did not correlate with the skilled listeners’ ratings of overall accent. These results suggest either that intonation measured in isolated sentences, as the POEC-S intonation subtest attempts to do, is not adequate to measure accent or that listeners are unable to assign a distinct intonation rating that is not influenced by the other components of accent.

Another interesting finding was that speakers’ POEC-S total scores and POEC-S verbal scores had an extremely strong positive correlation, suggesting that the auditory discrimination subtest score of the POEC-S does not account for much (i.e., 4%) of the total variance in total POEC-S score. However, in all correlations that were calculated in this study using the POEC-S total score and the POEC-S
The correlations were stronger with the POEC-S total score than with the POEC-S verbal score, indicating that the POEC-S total score has relatively stronger validity than the POEC-S verbal score does.

The best method for establishing the criterion validity of a given test is to compare it to an already existing standardized measure (McCaulley & Swisher, 1984). However, there are no standardized tests of accent available for comparison to the POEC-S. The test closest to being a criterion measure for the POEC-S is the TOEFL, which is a standardized test with documented validity that measures many aspects of English proficiency. POEC-S and TOEFL scores were positively correlated for the speakers in our study. This correlation, although significant, was less strong than the correlation between skilled listeners’ accent ratings and the POEC-S scores, and this makes sense. The skilled listeners’ accent ratings and the POEC-S scores are specifically related to accent, whereas the TOEFL score is a more general test of overall English proficiency, including both expressive and receptive skills.

Accent

Although our study was not intended to be a factor analysis of perceptual measures contributing to the perception of accent, our results do appear to shed some light on the evolving definition of accent. All parameters rated in the current study correlated with the speaker’s POEC-S total scores, so it appears that perceptions of articulation, intonation, naturalness, and estimated intelligibility contribute to perceptions of accent, as hypothesized by various authors (Gilbert, 1994; Morley, 1994; Pennington & Richards, 1986; Sikorski, 2005a, 2005b, 2005c; Wong, 1986). However, other parameters such as nonverbal language, fluency, or voice that were not included in this study may also play a role in accent.

Accent and Intelligibility

In our study, ratings of accent and estimated intelligibility were significantly positively correlated. These findings are similar to those of vanWijngaarden et al. (2002) and contrary to those of Munro and Derwing (1999). These contradictory findings may be due to the nature of the samples used, the types of measures used, or both. In terms of samples used, both our study and that of vanWijngaarden et al. used a heterogeneous sample of L2 speakers, sampling across a variety of L1s. Munro and Derwing’s study, on the other hand, used a relatively homogeneous sample of Mandarin speakers. Perhaps using a homogeneous sample allows one to sample across the entire possible spectrum of accentedness, whereas using a heterogeneous sample leads to a clustering of ratings that may not prove to be significantly correlated with intelligibility. Regarding measurements used, perhaps the differences are due to the different ways in which these studies chose to measure intelligibility. In ours and vanWijngaarden’s study, listeners used rating scales to estimate intelligibility; in Munro and Derwing’s study, listeners wrote down word-for-word what they heard. Both methods are reliable, and both methods have advantages and disadvantages (Munro, 1998; Watson & Schlauch, 2008; Yorkston & Beukelman, 1978). In our study, we chose to use estimates of intelligibility because we were asking listeners to rate multiple aspects of what they heard, and because estimates of intelligibility are commonly used in SLPs’ clinical practice. Finally, the fact that unskilled listeners were reliably able to rate the estimated intelligibility of the speakers but were not able to reliably rate accent attests to the fact that accent and estimated intelligibility are not identical parameters.

Accent and Naturalness

In this study, accent and naturalness ratings by unskilled listeners were not used in the analyses because their intrarater reliability was less than the criterion of 80%. This information is a finding in itself, indicating that when making perceptual ratings of accent and naturalness, the level of training of the listener plays a role. The level of listener
training/experience has been shown to be a critical variable in other studies as well (Berliner, 1994; Brundage, Bothe, Lengeling, & Evans, 2006; Cordes & Ingham, 1995). SLPs were considered to be skilled listeners due to their educational experience. The SLPs who participated in this study were able to rate all of the parameters reliably, but this was not the case for the unskilled listeners. The finding that unskilled listeners were unable to reliably rate naturalness or accent is contrary to some previously reported literature. Previous studies in which unskilled listeners rated the naturalness of speakers with various speech disorders resulted in intrarater reliability ranging from 80% to 90% within one scale value on a 9-point scale (Eadie & Doyle, 2002; Mackey et al., 1997; Martin et al., 1984). Mackey et al. (1997) also had unskilled listeners rate dialect on a 9-point scale and reached an intrarater reliability of 93%. The definition of and exclusionary criteria for dialect speakers used in that study were different from those used in the current study, so those factors may account for some of the inconsistencies. It is possible that raters are more reliably able to rate regional dialect speakers than foreign accent speakers, possibly due to increased familiarity with dialects or predictability of pronunciation differences. One major difference between the previous studies and the current study is that in the previous studies, raters were asked to rate the naturalness of speakers with documented speaking disorders, such as stuttering and tracheoesophageal speech.

**Accent, Articulation, and Intonation**

Although it remains unclear what the relative contributions of each of the components of accent have toward the perception of accent as a whole, vanWijngaarden et al. (2002) suggested that “clarity of articulation is the most important factor for the perception of overall accent strength” (p. 3,007). Our findings add preliminary support to the idea that articulation, more so than intonation, contributes to judgments about accented speech. In our study, correlations between skilled listeners’ accent ratings and speakers’ POEC-S *intonation* subtest scores were not significant, whereas correlations between skilled listeners’ accent ratings and speakers’ POEC-S *articulation* subtest scores were significant.

**Limitations**

One aspect of this study that limits generalization is that the accented individuals who provided speech samples had high TOEFL scores. The speaker samples used in the current study were heterogeneous for foreign language and number of years of experience speaking English. Therefore, the results of this study can only generalize to individuals with foreign accents who scored high on the TOEFL, but generalizations are not limited to a particular native language or experience.

Although perceptual measurements are essential to the field of speech-language pathology, they have inherent limitations and may be influenced by a myriad of factors. Factors related to the samples themselves include order of presentation, speech task (single words, reading, conversation), degree of accent, and type of accent (Brown & Docherty, 1995; Darley et al., 1969; Graham, 1950; Kent, Kent, Rosenbek, Vorperian, & Weismer, 1997; Poulton, 1989; Schmid & Yeni-Komshian, 1999; Shrivastav, Sapienza, & Nandur, 2005; Zeplin & Kent, 1996). We attempted to control for these factors by counterbalancing the order of sample presentation and limiting the samples heard to spontaneous speech. We decided to sample from a wide variety of L1s in order to allow generalization of our findings to a number of different L1s. It is impossible to control all of the factors that might contribute to ratings of accent when using spontaneous speech samples; however, spontaneous speech is the most ecologically valid type of sample for perceptual speech ratings (Coelho, 1998; Speaks et al., 1972).

Other factors that may influence perceptual ratings are related to the rating procedures: the number of parameters rated at once and the individual parameters that are rated (Bunton, Kent, Duffy, Rosenbek, & Kent, 2007; Kreiman & Gerratt, 1998; Miller, 1956). Some parameters have been rated more reliably than others (Zeplin & Kent, 1996), although it is important to note that accent and intelligibility can be measured reliably via interval scaling or percentage estimation (Southwood & Flege, 1999; Speaks et al., 1972). For our study, we used measurement methods that had been shown to be reliable and valid methods for use in perceptual studies.

Listener-related factors may also influence perceptual measures. These include familiarity with the topic, short-term memory, motivation, fatigue, interpretation of the rating scale, training, and experience (Duffy & Pisoni, 1992; Graham, 1950; Greene & Pisoni, 1988; Hammen, Yorkston, & Dowden, 1991; Pisoni & Luce, 1986; Platt, Andrews, Young, & Quinn, 1980; Shrivastav et al., 2005; Tjaden & Liss, 1992, 1995; Watson & Schlauch, 2008; Weismer & Martin, 1992; Yorkston & Beukelman, 1980; Zeplin & Kent, 1996; Zyski & Weisiger, 1987). We attempted to control for these factors by practicing with participants before the listening task, including familiarizing them with the rating scales used. We limited the amount of ratings to be completed so as not to tax short-term memory and allowed listeners to take brief breaks if they became fatigued. Speakers were limited to a closed set of possible topics in order to make samples similar but not identical. It is possible, however, that some listeners grew tired but did not tell us, or that the number of ratings required of each of 35 samples was too taxing. Additionally, 95% of the listeners who participated in this study were female. This sample is representative of the SLP population (ASHA, 2008); however, it is not representative of the general public. Therefore, any gender differences that may exist for unskilled listeners rating accent and its components were not measured in the current study.

**Future Directions**

First and foremost, more information is needed in order to determine whether the results found in the current study will generalize to other populations of English speakers with foreign accents. In particular, it would be beneficial
to use a sample of individuals representing a wide range of occupations, degree of accent, and English proficiency levels for a future study in order to determine whether the POEC-S has validity for the general population of speakers with accents.

Although the POEC is a widely used measure of accent and the POEC-S is convenient and has now been shown to be valid, many questions about the POEC-S remain. Test–retest reliability is another valuable feature of assessment tools, particularly when used to document treatment effectiveness. The POEC-S has two forms for this reason, but the current study did not examine the reliability of the POEC-S. The methodology used in this study could be replicated to determine the validity of other assessment tools (e.g., the full POEC, CPAFA). Furthermore, direct comparison studies of the available assessments could inform accent modification services as well as the definition of accent.

The current study also raised issues of reliability of perceptual ratings in unskilled listeners. These findings suggest that listener experience plays a role in certain perceptual ratings, similar to results found in other studies (Brundage et al., 2006; Platt et al., 1980; Zeplin & Kent, 1996; Zyski & Weisiger, 1987). Although several potential explanations for the low intrarater reliability were discussed earlier, it appears that more research is needed in this area. For example, it would be interesting to manipulate the degree to which samples differ from one another in order to quantify “just noticeable differences” between accented speech samples.

CONCLUSION

In summary, this study established the construct, criterion, and social validity of the POEC-S. Speakers’ POEC-S scores correlated with listeners’ perceptual ratings of accent, articulation, intonation, naturalness, and estimated intelligibility, implying that these parameters are all related in some way to the perception of accented speech. Interestingly, unskilled listeners were unable to rate accent or naturalness reliably, but skilled listeners were able to rate all parameters reliably. Speakers’ TOEFL scores also correlated with their POEC-S scores, but to a lesser degree than listeners’ perceptual ratings. These results have implications for SLPS’ clinical practice and the reliability of perceptual measures. SLPs have the training and knowledge to reliably rate accent and related parameters and are therefore appropriate service providers for accent modification. Clinicians should feel confident using the POEC-S with populations comparable to the one in this study. Further investigations of POEC-S, as well as other published measures of accent, are necessary if evidence-based practice guidelines are to be developed for accent modification services.

REFERENCES


164 CONTEMPORARY ISSUES IN COMMUNICATION SCIENCE AND DISORDERS • Volume 37 • 153–166 • Fall 2010


Schmid, P. M., & Yeni-Komshian, G. H. (1999). The effects of speaker accent and target predictability on perception of


Contact author: Shelley B. Brundage, George Washington University, Speech and Hearing Science Department, 2115 G Street NW, Suite 201, Washington, DC 20052. E-mail: Brundage@gwu.edu.