

Thompson, L. F., & Surface, E. A. (2006, May). *Assessment from Afar: Comparing telephonic and face-to-face foreign language speaking proficiency interviews*. Paper presented at the 21<sup>st</sup> annual meeting of the Society for Industrial and Organizational Psychology, Dallas, TX.

# **Assessment from Afar: Comparing Telephonic and Face-to-Face Foreign Language Speaking Proficiency Interviews**

Lori F. Thompson  
North Carolina State University

Eric A. Surface  
Surface, Ward, & Associates



**MAY 2006**

APPROVED FOR PUBLIC RELEASE;  
DISTRIBUTION UNLIMITED

# Copyright Notice

This document and its content is copyright ©1997-2010 of SWA Consulting Inc. All rights reserved.

Any redistribution or reproduction of part, or the entire document in any form is prohibited except for: (1) you may print or download to a local hard disk extracts for your personal and non-commercial use only, and (2) you may copy the content to individual third parties for their personal use, but only if you acknowledge the website and author(s) as the source of the material. You may not, except with our express written permission, distribute or commercially exploit the content, nor may you transmit it or store it on any other website or other form of electronic retrieval system.

Thompson, L. F., & Surface, E. A. (2006, May). *Assessment from afar: Comparing telephonic and face-to-face speaking proficiency interviews*. Paper presented at the 21st annual meeting of the Society for Industrial and Organizational Psychology, Dallas, TX.

## Assessment from Afar: A Field Study Comparing Telephonic and Face-to-Face Foreign Language Speaking Proficiency Interviews

Lori Foster Thompson  
North Carolina State University

Eric A. Surface  
Surface, Ward, & Associates

Today's ever-shrinking world is marked by a growing need for foreign language skills. This need is particularly pressing in the military, due to operations in Iraq, Afghanistan, and elsewhere. The demand for language-enabled personnel requires a valid mechanism for assessing language proficiency. Typically, an Oral Proficiency Interview (OPI) is used to evaluate speaking skills. Although technological advances and logistical considerations continue to move OPIs from face-to-face to technology-mediated formats, little is known about whether interview modality affects proficiency scores. The present study tested whether telephonic and face-to-face OPIs yield different ratings. We used a variety of analytic approaches to examine two samples of Special Operations Forces personnel ( $N=32$ ;  $N=52$ ) who completed a Defense Language Institute OPI in person or telephonically. Results supported the study hypothesis: telephonic examinees received less favorable scores than their face-to-face counterparts even after controlling for interview year, language difficulty, cognitive ability, and language aptitude.

The contemporary work world is marked by a growing need for foreign language skills. "In today's workplace, many companies, agencies, corporations, and other institutions are experiencing ever-increasing demands to hire personnel with language skills" (Swender, 2003, p. 524). This need is particularly pressing in the military, due to operations in Iraq, Afghanistan, and elsewhere.

The demand for language-enabled personnel requires a valid mechanism for assessing language proficiency. Typically, an Oral Proficiency Interview (OPI) is used

to evaluate speaking skills in military, governmental, and private-sector organizations. Although technological advances and logistical considerations continue to move speaking proficiency interviews from face-to-face to remote, technology-mediated formats, little is known about whether the interview modality affects proficiency scores. The lack of research on this topic represents a significant gap in the literature, particularly because foreign language proficiency is likely to play an increasingly important role in both military and private sector organizations during the

*Note.* Please do not cite without permission from the authors.

days to come (e.g., Cascio, 2003; Rovira, 2003). The purpose of the present study is to test whether United States (U.S.) Special Operations Forces Soldiers tested via a telephonic OPI receive lower scores than those tested face-to-face. Given that the OPI is used as a language training graduation standard in military organizations (e.g., the Defense Language Institute) and is the official measure of speaking proficiency for the U.S. military, modality differences can have real consequences for units and their Soldiers, Sailors, Airmen, and Marines.

### *The Importance of Foreign Language Proficiency and Assessment*

Foreign language usage and the assessment of oral proficiency continue to be major issues in the U.S. military. A large number of military occupational specialties require intensive use of foreign languages (Silva & White, 1993). Recent U.S. operations in Iraq, Afghanistan, the Horn of Africa, and Colombia have caused military leaders to intensify their focus on the foreign language capabilities of U.S. troops (e.g., Rovira, 2003). Operations within and beyond these countries have increased the need for language-enabled operators and linguists whose proficiency must be assessed and documented for developmental and administrative purposes.

The needs for language proficiency and assessment are especially great within the Special Operations Forces community (Surface, Dierdorff, & Donnelly, 2004; Surface, Poncheri, Lemmond, & Shetye, 2005), which accounts for a large percentage of the language billets in the Department of Defense. Soldiers, Airmen, Sailors, and Marines working in Special Operations Forces units conduct complex and important missions (e.g., Unconventional Warfare) around the globe. Consequently, these units often require teams of operators who are collectively proficient in multiple languages.

Language testing is a noteworthy concern not only for U.S. military forces, but also for military teams in Central and Eastern Europe (Green & Wall, 2005). The issue of language proficiency and assessment extends to non-military organizations as well. In 2004, an unprecedented meeting of leaders from U.S. government, academia, industry, and language associations was held to discuss the critical need for enhanced national foreign language capabilities. The result of this meeting was a documented “Call to Action for National Foreign Language Capabilities,” (United States Department of Defense, 2005), which laid out a series of recommendations to address the nation’s foreign language deficiencies. Among these recommendations was the creation of a council which would provide guidance on the standardized assessment of achievement and proficiency in foreign languages. According to the 2005 Call to Action for National Foreign Language Capabilities, “... no one sector – government, industry, or academia – has all of the needs for language and cultural competency, or all of the solutions” (United States Department of Defense, 2005, p. ii).

In the private sector, the importance of oral language assessment is escalating, in part because labor mobility is on the rise. Immigrant and refugee flows have become an increasingly entrenched component of the current global situation (Chan, Schmitt, Jennings, & Sheppard, 1999; McNamara, 1998). In some areas of the world, effective functioning in nearly all jobs requires proficiency in the country’s native language (Chan et al., 1999). To counter language deficiencies in the workplace, some organizations have begun to offer foreign language training (Hammers, 2005). According to Noe (2005), approximately 20-30% of organizations have a budget allocated for this function.

Worker mobility and its language implications have also increased due to expatriate assignments by multinational corporations. Cascio (2003) deems globalization a defining characteristic of the twenty-first century and suggests that the need for expatriate language training is one of its consequences. Both within the U.S. and elsewhere, language skills can affect expatriate adjustment to the host country (Chao & Sun, 1997; Takeuchi, Yun, & Russell, 2002) by influencing the accuracy of their expectations prior to the assignment (Caligiuri, Phillips, Lazarova, Tarique, & Burgi, 2001) and by shaping the relationships formed during the assignment (Caligiuri & Lazarova, 2002).

Finally, large companies have been moving jobs abroad, and this offshoring trend is expected to continue (Chittum, 2004; Pristin, 2003; Smith & Frangos, 2004). This movement has implications for language usage and assessment. Call centers are growing rapidly, both in terms of the number of people employed and the increasing size of the sector (Callaghan & Thompson, 2002). Meanwhile, companies located in the U.S. and the United Kingdom continue to outsource their call centers to countries such as India and the Philippines (Pristin, 2003; Vina & Mudd, 2003). As a result, many call centers employ personnel whose native language differs from the language spoken by the majority of the customers they serve. An article by Flynn (2003) addresses one effect of this trend – training workers abroad to be sensitive to the cultures of customers located in the U.S. and elsewhere. At the most basic level, this sensitivity requires language skills.

#### *The Assessment of Oral Proficiency in a Foreign Language*

The increasing need for foreign language proficiency draws attention to the methods used to assess language skills. The OPI is a

procedure commonly used to evaluate spoken language proficiency. The OPI is a semi-structured, 15-30 minute conversation between one or more trained interviewers and an examinee. It is designed to push examinees to the limits of their spoken competence, thereby eliciting a profile of oral proficiency (Halleck, 1995; Malone, 2003). The OPI follows an established protocol; however, there is no script or prescribed set of questions (Swender, 2003). It tends to be more like a formal interview than a natural conversation (Johnson, 2001, as cited in Malone, 2003). With common versions of the OPI, interview performance is scored on the following 11-point scale: 0, 0+, 1, 1+, 2, 2+, 3, 3+, 4, 4+, and 5. Zero represents no functional ability, and 5 represents oral proficiency equivalent to an educated native speaker (Herzog, 2003). Two certified raters, at least one of whom serves as an interviewer during the exam, use scoring guidelines to independently evaluate the examinee's performance.

Since World War II, the U.S. government has used the OPI to assess the language skills of American personnel working abroad (Halleck, 1995). OPIs are used by the Central Intelligence Agency, the Peace Corps, the Defense Language Institute, and other military, educational, and private sector institutions. The Defense Language Institute alone conducts over 3,000 OPIs every year. This large-scale effort requires more than 300 certified OPI testers proficient in dozens of different languages, dialects, and variants (Herzog, 2003).

OPI assessments often carry significant consequences. The OPI is administered to evaluate faculty applicants and other government employees (Herzog, 2003). It is also used to assess students completing language training courses. In the future, military personnel (whose foreign language proficiency pay requires a regular

assessment of proficiency in a target language) may be allowed to demonstrate language proficiency via the OPI. As a result, OPI scores may affect the income of many who take the exam.

Recent investigations have produced preliminary support of the OPI construct (Dandonoli & Henning, 1990; Glisan & Foltz, 1998), and the OPI has been found to be reliable (Adams, 1978; Bachman & Palmer, 1981; Clark, 1986; Magnan, 1986; Surface & Dierdorff, 2003; Thompson, 1995). Nevertheless, OPI evaluations are not purely objective. While scoring guidelines aid interrater reliability, evaluations involve a degree of personal judgment. The result is that OPI testers draw upon their built-in, idiosyncratic rating criterion schemes when evaluating examinees' performance (Chalhoub-Deville & Fulcher, 2003).

#### *The Use of Non-Face-To-Face OPI Assessment Modalities*

There is considerable interest in using alternative OPI modalities when the administration of a face-to-face interview is either impractical or impossible (Jackson, 1999; Stansfield & Kenyon, 1992). At present, existing alternatives to the face-to-face format include interviews conducted via telephone, videoconferencing, and a tape-mediated procedure (Stansfield & Kenyon, 1992) which removes the human interviewer from the process and instead delivers interview instructions and questions via a master tape while asking examinees to record their responses on separate tapes. One of the most recent alternatives under development is an Internet-delivered, semi-direct version of the OPI called the ACTFL OPIc<sup>®</sup>. This assessment uses computer technology to elicit and collect a ratable sample of speech from the examinee, eliminating the need for the interviewer and allowing the sample to be scored by certified raters located off site (Surface, Poncheri, &

Bhavsar, 2006).

At present, telephonic assessment is probably the most practical and widely used alternative to the face-to-face OPI. Telephonic oral proficiency assessments allow for more flexible scheduling and cost-effective pricing when compared to the face-to-face alternative (Swender, 2003). Relative to a tape-mediated procedure, a telephonic exam more closely approximates the face-to-face conversation. Unlike a videoconference-based exam, the technology required for a telephonic exam is familiar and readily available to most. The telephonic modality is attractive because assessments can be arranged anywhere in the world and many examinees can be tested within a short period of time (Swender, 2003).

According to Anderson (2003), equivalence cannot be taken for granted in alternative administrations of assessment methods, even when the face-to-face alternative is as basic in nature as telephone-based interview procedures. Although the use of non-face-to-face OPI modalities is common, the effects of modality are not well understood and need to be examined (Swender, 2003).

The dearth of literature addressing modality effects on the OPI necessitates a careful inspection of other bodies of work outside of the language testing domain. The literature considering modality effects during the employment interview process provides insights regarding the manner in which telephonic and face-to-face OPIs may differ. Similar to an OPI, the employment interview is a semi-structured conversation that allows an interviewer to evaluate an examinee's knowledge, skills, etc. Insights from the employment interviewing domain are considered below, along with supplementary findings from the language literature.

### *Modality Preferences*

One way to touch on the possibility of a modality effect is to examine whether people prefer one interview format to another. Studies outside of the language domain have looked at reactions to assessment methods administered across different modalities. In 2003, Paronto, Bauer and Truxillo (as cited in Anderson, 2003) compared applicants' reactions to a telephonic versus a face-to-face screening test. The face-to-face assessment was viewed more positively than its telephonic alternative.

An unpublished technical report by Jackson (1999) examined cross-mode reactions to the OPI in particular. Examinees' questionnaire responses revealed a preference for the face-to-face compared to the telephonic interview. More than 70% of Jackson's (1999) participants considered the telephonic OPI inferior to the face-to-face alternative. During the face-to-face interview, 85% of the examinees viewed the test as fair. Conversely, only 61% of the examinees perceived the telephonic test as fair. When asked which modality they would prefer for future testing, approximately 68% of the respondents, who had taken both face-to-face and telephonic OPIs, replied "Face-to-face." By comparison, 23% preferred the telephonic modality and 8% had no preference. These results suggest that examinees may perceive incongruity between face-to-face and telephonic interviewing.

### *Potential Differences Between Face-to-Face and Telephonic Modalities*

Examinees' preference for the face-to-face modality is perhaps unsurprising in light of the differences between face-to-face and telephonic interaction. As described next, there are several reasons to believe that the telephonic modality may put

interviewees at a relative disadvantage when compared to their face-to-face counterparts. This presumed disadvantage stems from four factors: communication processes; the interview content; examinee performance; and interviewer scoring. Although an empirical examination of these factors is beyond the scope of the present study, each issue is described below as justification for our forthcoming hypothesis that the telephonic modality puts OPI examinees at a relative disadvantage when they are compared to those tested by a face-to-face interviewer.

*Communication processes.* Inherent differences between face-to-face and telephonic interactions raise concerns about the possibility of a modality effect on OPI scores. Telephone and face-to-face communication are qualitatively different from each other. For starters, the physical limitations of telephone communication may restrict the full range of sounds needed to clearly express oneself in a foreign language (Swender, 2003). OPI research by Jackson (1999) reinforces this point. An interviewer participating in Jackson's research reportedly noticed the existence of differences in intonation patterns between the face-to-face and telephonic modalities. Another interviewer was said to have expressed discomfort using the telephonic interview to assess proficiency because nuances of voice and intonation seemed "lost" over the telephone (Jackson, 1999). Such nuances might be particularly critical for the assessment of "tonal languages."

Sound quality is not the only difference between face-to-face and telephonic communication. Telephone conversation is marked by "cuelessness" because it is constrained to sounds (Drummond & Hopper, 1991). Compared to its face-to-face alternative, the telephone is considered a lower bandwidth medium because it filters visual information and transmits fewer cues.

The lower the bandwidth, the less information about traits and behaviors can be exchanged (Straus, Miles, & Levesque, 2001).

During telephonic interviews, examinees and interviewers have no eye contact, and “they cannot monitor nonverbal information such as body posture, head movements, hand and facial gestures” (Silvester, Anderson, Haddleton, Cunningham-Snell, & Gibb, 2000, p. 17). Notably, a lack of visual cues prevents the transmission of certain forms of backchannel feedback. While some forms of this feedback are verbal and can be conveyed over the telephone (e.g., pauses and utterances such as “uh-huh”), many are visual and non-verbal (e.g., eye contact, facial expressions, head nods and shakes). According to Straus et al. (2001), backchannel feedback is important because it facilitates listener understanding and coordinates conversations. It allows conversation participants to monitor each other’s behavior and to better control and correct their own performance (Silvester et al., 2000).

Limited opportunities for backchannel feedback raise concerns that speech may be less synchronized during telephonic compared with face-to-face interviews (Silvester et al., 2000). Silvester and Anderson (2003) report that telephone conversations contain fewer interruptions, longer utterances, and more questions than face-to-face conversations. A lack of interruptions, while generally associated with politeness, is not necessarily a constructive communication feature. Indeed, it may signal the absence of interactive listening. Interactive listening requires the listener to take an active role by interacting with a speaker, requesting clarification, or providing feedback in order to ensure successful communication (Vandergrift, 1997). Interruptions from listeners can therefore serve valuable feedback functions.

Rutter and Stephenson (1977) “suggest that interruptions occur more freely in face-to-face communication because the visual channel allows the additional communication of nonverbal signals which help to maintain the interaction and prevent breakdown” (Silvester et al., 2000, p. 17). Thus, the lack of interruptions characterizing telephone communication can result in asynchronous speech, which could lead to difficult and disjointed interviews (Silvester et al., 2000).

The non-language, employment interviewing literature acknowledges that uncoordinated and disjointed communication may adversely affect perceptions of applicants’ verbal fluency and negatively impact interviewers’ evaluations (Straus et al., 2001). Language researchers agree. “Consider, for example the task of assessing interactive comprehension face to face versus over the phone. It is likely to be more difficult to assess that factor in the latter modality due to the lack of nonverbal visual indications of comprehension on the part of examinees” (Jackson, 1999, p. v). Facial expressions and other visual, non-verbal gestures add context to speech. These context cues may help face-to-face examinees decipher the interviewer’s speech. As well, the ability to receive nonverbal cues and gestures can fill gaps in verbal communication when interviewers are evaluating examinees (Swender, 2003).

*The interview content.* For better or worse, the OPI tester is intimately implicated in the construction of examinee proficiency (Brown, 2003). Different OPI ratings on different occasions are partially a function of the questions and tasks the interviewer chooses to include, which drive the grammatical structures and lexicon needed for a correct response on the part of the examinee (Jackson, 1999). The telephonic communication characteristics

discussed above may cause OPI testers to compensate for modality limitations by simplifying the content of the telephonic interview. This could lead to relatively unsophisticated examinee responses and low scores.

Anecdotal evidence supports the claim that the OPI medium can impact interview content. According to Jackson's (1999) research, "Interaction between test modality and content was evidenced by one Russian tester who wrote: '...I had tendency to simplify my language during telephonic in order to avoid misunderstanding due to means of communication.' The same tester reported that she felt freer to ask certain questions face to face because she didn't worry that the examinee would be confused" (p. 22).

*Examinee performance.* The OPI modality may affect the way the examinee performs due to communication breakdowns, uneasiness with the interview modality, and a lack of rapport with the interviewer. The first of these three issues was alluded to earlier: the relative cluelessness characterizing telephone communication can present telephonic examinees with challenges and communication gaps not present during a face-to-face interview. This alone may cause telephonic examinees to perform worse than their face-to-face counterparts.

In addition, modality effects on the examinee's mental state during the interview may shape performance. OPI examinees' mental state is influenced by anxiety, modality preferences, and the like (Jackson, 1999). As noted, people generally prefer the face-to-face OPI modality to its telephonic counterpart. One would therefore expect that a more positive mental state will characterize examinees taking a face-to-face compared to a telephonic OPI. Said another way, telephonic examinees may enter the interview with more a priori uneasiness

about the process. People tend to make inferior impressions to the extent that they feel less comfortable during an interview (Straus et al., 2001). Hence, examinees' modality apprehensions can adversely affect performance.

Finally, difficulties establishing rapport may put telephonic examinees at a relative disadvantage. Telephone conversations have an abrupt beginning and end (Drummond & Hopper, 1991), whereas face-to-face communication allows people to ease into an interview. During conversations, emotions and reactions are expressed by visual cues such as nodding, smiling, eye contact, and body position. These expressions, which are absent during telephone interaction, influence the development of rapport and social relationships (Straus et al., 2001).

Rapport between the interviewer and examinee can influence performance. People vary their language when interacting with familiar and unfamiliar others. Indeed, past language testing research has shown an acquaintanceship effect, with examinees achieving higher performance levels when communicating with a friend compared to a stranger (O'Sullivan, 2002).

In short, telephonic examinees may perform worse than their face-to-face counterparts because the remote medium offers fewer opportunities for relationship building and rapport with interviewer. Anecdotal evidence from the language literature supports the assertion that variance in OPI ratings is partially attributed to the establishment of rapport, which is fueled by nonverbal interactions (Jackson, 1999).

*Interviewer scoring.* Interviewer reactions to information presented by examinees may also differ across face-to-face and telephonic modes of presentation (Anderson, 2003). As such, the modality may affect the way the interviewer evaluates performance. Specifically, modality may have a direct effect on scores, above and

beyond any indirect effects mediated by changes in tester and examinee behaviors.

As mentioned earlier, telephonic interviews offer fewer opportunities than their face-to-face analogues for relationship building and rapport between the interviewer and the examinee. Irrespective of performance, this rapport and relationship development may affect the way the examiner scores the interview. In effect, cue deprivation can increase the psychological distance between the interviewer and interviewee, leading to harsher ratings of the interviewee (Silvester et al., 2000).

#### *Research Comparing Face-to-Face and Telephonic Interviews*

Of the few published studies comparing telephonic and face-to-face interviews, most have occurred within the employment interviewing domain. Straus et al. (2001) examined modality effects during mock employment interviews conducted with college students. Notably, there were no significant differences between the conversational fluency ratings assigned by the face-to-face versus telephonic interviewers who participated in this study. However, the examinees reported that conversations flowed somewhat more smoothly in face-to-face compared to the telephonic conversations (Straus et al., 2001). In terms of performance, the interview modality did not affect the scores assigned to relatively attractive interviewees. Relatively unattractive examinees, however, were scored higher in the telephonic versus the face-to-face medium. Straus et al. (2001) concluded their study by cautioning that students' responses to interviews conducted solely for the sake of research may not generalize to actual interview situations.

Silvester et al. (2000) and Silvester and Anderson (2003) investigated the equivalence of competency based, 30-

minute, semi-structured interviews conducted face-to-face and telephonically. Both of these studies examined real-world interviews with authentic employment consequences. In both investigations, interviewers rated applicants lower in the telephonic compared to the face-to-face condition.

Moving beyond the employment interviewing domain, two studies have looked at the telephone's effects on OPI performance. First, a technical report produced by Jackson (1999) described a study examining the possibility of modality effects on OPI scores. Sixty-seven volunteers completed both a face-to-face and a telephonic interview in Russian, with a different pair of interviewers assessing the examinees during the two interviews. Results showed no significant difference between the face-to-face and the telephonic conditions.

Jackson's (1999) face-to-face interviews were videotaped, which may have added an atypical degree of tension to those exams. Moreover, the interviews examined in this study were not for record – that is, they had no real consequences. Jackson (1999) emphasized this factor as a study limitation. The employment interviewing literature highlights the distinction between interviews studied for their own sake and those conducted in field settings with authentic selection decisions (Silvester et al., 2000). Studying real-world interviews with genuine consequences is said to be of paramount importance (Anderson, 2003).

Swender's (2003) OPI research examined the effects of modality on OPIs conducted for record in the sense that official OPI ratings and certificates were awarded upon study completion. Thirty-four individuals completed Spanish OPIs in both a face-to-face and a telephonic setting. The testers in this study were highly trained individuals who conducted hundreds of

interviews each year. Standard OPI procedures were adjusted to accommodate the experimental nature of the assessment. For example, the same tester conducted both the face-to-face and the telephonic interview for a given examinee. Moreover, in contrast with the typical OPI operating procedures, interviewers did not score the examinees' performance.

The results of Swender's (2003) study showed no significant differences based on interview modality. The study design raised some questions that can be clarified by follow-up research. For instance, the interviewers in Swender's (2003) study found it difficult to deliver comparable but different interviews to the same individual across the two modality types. During the second interview, they "reported that it required a conscious effort on the tester's part to avoid topics already explored in the first interview" (Swender, 2003, p. 522).

As discussed earlier, the hypothesized modality effect on OPI scores may be partially driven by differences in the way testers conduct face-to-face and telephonic interviews. In Swender's (2003) research, the same interviewers examined the interviewee within both modalities. In all likelihood, the interviewers made a concerted effort to offer an equivalent experience across modalities. This concerted effort may not exist during everyday OPIs where interviewers are not participating in research study examining modality effects.

Other features of Swender's (2003) design may have further prevented differences between face-to-face and telephonic OPIs from surfacing. During everyday OPIs, difficulties establishing rapport and psychological distance may cause interviewers to inadvertently assign harsher ratings to telephonic examinees. Because Swender's (2003) OPI procedure was adjusted so that interviewers did not

score examinees, this potential problem was all but eliminated.

Clearly, there is a need to replicate existing research on the OPI (Malone, 2003). Indeed, Swender (2003) concluded her study by calling for further investigation of the potential effects of modality on OPI scores. Follow-up research comparing telephonic and face-to-face OPIs conducted in naturalistic, field environments would be informative. In this vein, the present study tests the following hypothesis with data from a real-world testing situation, in which OPI performance had legitimate consequences.

*Hypothesis:* Individuals completing OPIs telephonically will receive lower ratings than their face-to-face counterparts.

## Method

### *Participants*

Two different samples were separately examined in this study. The first sample consisted of 32 military personnel within the Special Operations Forces community who attended a foreign language training program between 1997 and 2001 at the Defense Language Institute. This training lasted between 25 and 63 weeks, depending on the difficulty level of the language being taught. The Defense Language Institute is the official language authority for the U.S. military and trains mostly personnel who are in the military intelligence and foreign area officer career fields. While the Special Operations Forces community as a whole sends very few personnel to the Defense Language Institute, some exceptions requiring high levels of language proficiency are trained there. For Special Operations Forces operators, attendance at the Defense Language Institute is voluntary and dependent on the availability of seats in the target training program.

The second sample consisted of 52 Special Operations Forces personnel who attended a foreign language training program between 1996 and 2003 at the U.S. Army John F. Kennedy Special Warfare Center and School (USAJFKSWCS). This training, which lasted between 18 and 24 weeks depending on the difficulty level of the language being taught, is commonly completed by Special Operations Forces operators who are required to establish a minimum proficiency in a foreign language.

### *Design and Procedure*

Individuals at both the Defense Language Institute and USAJFKSWCS are required to complete an OPI for record at the conclusion of their language training. Examinees' OPI scores are recorded and used for several purposes. For example, they serve as a criterion measure in research and evaluation studies for new courses. In addition, they are used as a graduation requirement for trainees learning relatively uncommon or low-density languages (e.g., Urdu) for which no alternate final exam has been developed. The OPI is also linked to the Defense Language Institute's speaking proficiency graduation standard. As well, the OPI has been used sporadically as a surrogate graduation criterion at USAJFKSWCS. During some periods, USAJFKSWCS trainees who have failed to achieve the graduation standard on either the listening or reading component of their final exam have been allowed to pass the course by replacing a substandard final exam score with an acceptable OPI score.

The same standardized protocol is used to administer and score OPIs at the Defense Language Institute and USAJFKSWCS. Two raters, at least one of whom serves as an interviewer during the exam, independently evaluate each examinee's performance. The OPI is administered either face-to-face or telephonically, depending on

a variety of factors. For example, telephonic interviews are a viable alternative when examinees leave the training site before completing an OPI. This can occur when scheduling constraints prevent a large number of examinees from being interviewed within a short period of time. In these instances, the examinee may be allowed to return home and then complete the assessment over the telephone. In addition, face-to-face interviewers are not always available on site for relatively uncommon languages. In this case, the telephone provides a mechanism for obtaining access to off-site certified OPI testers.

The OPI and other data generated by the Defense Language Institute and USAJFKSWCS examinees investigated in the present study were collected from archives provided by the Defense Manpower Data Center, which receives its data from the Defense Language Institute, military testing centers, and other military organizations. The Defense Manpower Data Center maintains this database for archival purposes (e.g., future use and historical record).

USAJFKSWCS also stores data generated by personnel who attend USAJFKSWCS language training. Data obtained from USAJFKSWCS archives were used in this study to add cases that were missing from the Defense Manpower Data Center data set and to fill in missing values within incomplete Defense Manpower Data Center cases. While the incorporation of the USAJFKSWCS data set reduced the problem of missing data, this problem was not altogether eliminated. To some extent, the archival nature of the data rendered incomplete cases inevitable.

### *Measured Variables*

*Predictor.* The main independent or predictor variable of interest was interview

modality. The database indicated the modality used to interview Defense Language Institute and USAJFKSWCS examinees. Face-to-face interviews were coded as “1,” and telephonic interviews were coded as “2.”

*Criterion.* The main dependent or criterion variable of interest was examinees’ OPI scores. There are several versions of the OPI in existence. The Defense Language Institute’s version of the OPI was used for this study. As such, the final OPI score included in the database was provided by two raters who independently assessed each examinee’s speaking proficiency based on what is known as the Interagency Language Roundtable scale. In the event of a major disagreement, a third rater provided an assessment as well, with the examinee’s final score representing the rating assigned by two of the three evaluators.

The Interagency Language Roundtable scale is ordinal and ranges from 0 (representing no functional ability) to 5 (representing a proficiency level comparable to that of a well-educated native speaker). The following 11 points comprise this scale: 0, 0+, 1, 1+, 2, 2+, 3, 3+, 4, 4+, and 5. For analytic purposes, these 11 scale points were recoded into sequential integers ranging from 0 to 10. A score of 0 from the original scale remained a 0; a score of 0+ was changed to 1; a score of 1 was changed to 2; and so forth.

The OPI scores used in the Defense Language Institute study sample ranged from 0+ to 3 with 1+ being the mode. The OPI scores in the USAJFKSWCS sample ranged from 0 to 3 with a mode of 1.

*Covariates / Control Variables.* Several covariates or control variables were included in this research. We controlled for the year in which the examinee was interviewed, the difficulty level of the language being tested, the examinee’s cognitive ability, and the examinee’s language aptitude. These

covariates were chosen because of their potential association with the criterion.

The year in which the interview occurred was included because the year 2000 was marked by a slight revision in the protocol for conducting the version of the OPI used in this study. Language difficulty was also important to consider. Four standard categories are used to describe the difficulty level of learning various languages for native English speakers. Category I includes languages such as French and Spanish, Category II includes languages such as German and Malay, Category III includes languages such as Russian and Greek, and Category IV includes the most difficult languages for native English speakers (e.g., Arabic and Chinese). Each participant received a score of 1-4 based on the pre-determined difficulty category of the language that was tested.

Cognitive ability was measured via the Armed Forces Qualification Test (AFQT), which participants completed prior to language training. The AFQT assesses mathematical reasoning, mathematical computation, vocabulary, and reading comprehension. AFQT scores, which are reported as percentiles, correlate well with standardized tests of intelligence (Orme, Brehm, & Ree, 2001).

Lastly, the Defense Language Aptitude Battery (DLAB) was used to measure language aptitude. This is a multiple choice test that requires examinees to learn and use an artificial language. It was devised to assess the most important abilities needed to learn a second language, such as the ability to process auditory phonetic material. The DLAB has been shown to contribute significant incremental validity to the prediction of language proficiency, above and beyond measures of general aptitude (*g*) (Silva & White, 1993).

## Results

Collectively, past research has analyzed the OPI with a combination of nonparametric (Jackson, 1999; Swender, 2003) and parametric (Glisan & Foltz, 1998) statistics. The present study follows suit, relying heavily on parametric statistics but presenting nonparametric equivalents when warranted and possible.

### *Sample 1: Defense Language Institute Study*

Interview modality and OPI scores were available for 32 of the Defense Language Institute trainees in the database. Sixteen (50%) of these individuals completed a face-to-face OPI and the remaining 16 (50%) completed the OPI telephonically. At first glance, participants who were interviewed in a face-to-face setting ( $M = 3.44$ ,  $SD = 0.81$ ,  $Median = 3.00$ ) appeared to receive higher OPI scores than those interviewed telephonically ( $M = 2.25$ ,  $SD = 1.65$ ,  $Median = 2.00$ ).

Table 1 shows the correlations among the study variables, thereby providing a preliminary view of the relationship between interview modality and OPI ratings. An examination of the modality column in Table 1 reveals some noteworthy associations. First, modality was confounded with language difficulty, with easier languages being tested in the telephonic format. Second, telephonic examinees received significantly lower OPI scores than face-to-face examinees ( $r = -.43$ ,  $N = 32$ ,  $p < .05$ ). More than 18% of the variance in OPI scores was accounted for by modality.

Next, an analysis of covariance (ANCOVA) was conducted to compare the face-to-face and telephonic OPI averages that would have occurred if all participants had been equal in terms of test date, language difficulty level, cognitive ability, and language aptitude. As the results in

Table 2 indicate, the ANCOVA sample size was reduced due to incomplete cases (where participants had scores on some but not all of the 6 variables of interest).

Together, modality and the covariates accounted for 78% of the variance in interview scores. Moreover, the OPI scores assigned to face-to-face examinees ( $M = 3.67$ ,  $SD = 0.82$ ) were significantly more favorable than the scores assigned to their telephonic counterparts ( $M = 1.55$ ,  $SD = 0.52$ ) even after removing any effects the covariates included in this study may have had. As shown in Table 2, the interview modality's effect size was substantial; consequently, modality emerged as a highly significant predictor despite the reduced power associated with the reduced sample size.

### *Sample 2: USAJFKSWCS Study*

The initial USAJFKSWCS database contained more than 2000 cases. However, only a limited number of these entries were ultimately usable. Many of the cases were dropped due to incomplete or missing data. Others were eliminated as part of a matching process, described next, which was conducted as part of paired samples design used to control for the effects of the four variables examined as covariates in the Defense Language Institute study (i.e., test year, language difficulty, cognitive ability, and language aptitude).

USAJFKSWCS trainees typically complete their OPIs telephonically; thus, the vast majority of the OPI scores in the database came from telephonic examinations. Nevertheless, the final database included face-to-face OPI scores, along with values for most or all of the other key variables of interest, for 26 participants. The large number of telephonic examinees in the data set provided an opportunity to equalize the face-to-face and telephonic groupings in terms of test year, language

difficulty, cognitive ability, and language aptitude. An ex post facto design was therefore employed wherein each of the 26 face-to-face participants was matched to a telephonic counterpart with regard to the four control variables of interest.

This pairing process began by searching for a telephonic case that matched each face-to-face case in terms of language difficulty, cognitive ability, and language aptitude. When more than one telephonic participant matched a face-to-face case on all three variables, all telephonic matches were listed and one telephonic participant was chosen on the basis of test date (to guard against a date/modality confound). When two or more of these telephonic finalists were equally close to the date in which the face-to-face participant was interviewed, then one of the telephonic finalists was chosen at random.

When exact matches were not found for all three variables, cases were first matched by language difficulty. Next, the absolute difference between the face-to-face and telephonic participants' cognitive ability scores was computed, as was the absolute difference between face-to-face and telephonic participants' language aptitude scores. These two absolute difference scores were summed, and the telephonic candidate with the lowest absolute difference was matched to the face-to-face participant under consideration. In the event of a tie (i.e., if two or more telephonic candidates matched the face-to-face participant equally well), then all of the ties were listed and one telephonic participant was chosen on the basis of test date. When two or more such telephonic finalists were equally close to the face-to-face participant on the basis of test date, then one of the finalists was chosen at random.

Table 3 provides the correlations among the study variables within the final sample of 52. The characteristics of the face-to-face and telephonic groups, respectively, are

shown in Table 4. A paired samples t-test was conducted to determine whether the two groups, which were virtually identical with respect to each variable except modality, received significantly different interview scores. The results revealed a highly significant difference ( $t(25) = 5.31, p < .001$ ).

Two nonparametric equivalents of the paired t-test were also conducted. The Wilcoxon Matched-Pair Signed-Ranks Test revealed significantly ( $p < .001$ ) lower OPI scores for telephonic compared to face-to-face examinees. Similarly, the Sign Test also demonstrated a significant ( $p < .001$ ) difference between the two groups.

Lastly, the type of ANCOVA analysis conducted on the Defense Language Institute sample was replicated on the 52 USAJFKSWCS cases (26 face-to-face and 26 telephonic). While the paired samples tests presented above are arguably more appropriate analyses for these data, the ANCOVA is included to facilitate comparisons between the USAJFKSWCS and Defense Language Institute results.

As indicated in Table 5, the ANCOVA reduced the sample size to 30 due to incomplete cases. Together, modality and the covariates accounted for 84% of the variance in interview scores. After removing any effects that the covariates included in this study may have had, the OPI scores assigned to face-to-face examinees ( $M = 3.56, SD = 0.88$ ) were significantly more favorable than the scores assigned to telephonic examinees ( $M = 1.76, SD = 0.77$ ).

## Discussion

The present study provides an important contribution to the literature by addressing a research need recently articulated by Green and Wall (2005): "There appears to be little literature available – either descriptive or research-related – on language testing in the

military. This form of specific purposes assessment affects both military personnel and civilians working within the military structure in terms of posting, promotion and remuneration, and it could be argued that it has serious social consequences if not carried out professionally and to the highest standard” (p. 379).

To address this need, we used a variety of parametric and nonparametric approaches to investigate modality effects on OPI scores assigned to Special Operations Forces personnel tested at two different locations. Each sample and analysis led to the same conclusion, which supported the study hypothesis: telephonic examinees received significantly less favorable scores than their face-to-face counterparts. Even two groups of individuals who were virtually identical with respect to every study variable except modality received significantly different interview scores. The implication is that completing an OPI telephonically, rather than face-to-face, may put an examinee at a relative disadvantage.

#### *Study Limitations and Strengths*

These results should be interpreted in the context of some noteworthy limitations. First, participants were not randomly assigned to interview modality. Given the archival nature of the data, it is impossible to know for certain why some examinees were interviewed in the face-to-face mode while others were interviewed telephonically. Thus, it is conceivable that factors other than modality, which were not measured in this study, caused the difference between the face-to-face and telephonic participants' scores.

Second, the rotation of military personnel every 18 to 24 months posed challenges in this study. This loss of institutional memory prevents a conclusive confirmation of the exact testing conditions employed during the collection of the OPI

data used in this study.

Third, a number of cases were eliminated because of incomplete data. The modality predictor of interest had a substantial effect and therefore produced significant results despite the power loss stemming from this sample size reduction. Nevertheless, it is important to question the reason why the study variables were not consistently maintained in the database. If data were omitted for non-random reasons related to the research issue under investigation, there would be cause for concern.

Its limitations notwithstanding, this investigation has several strengths worth mentioning. This is the first study to examine modality effects using the standard OPI protocol in a naturalistic, field environment. Unlike some past research, the face-to-face interviews examined in this study were not videotaped for research purposes. Rather, the face-to-face interviewers and participants studied here functioned under normal OPI circumstances, without the added anxiety of being research participants captured on video.

In contrast to some prior work, the interviews examined in this study were for record. Therefore, this research offers a glimpse at realistic testing results, which reflect the dynamics of testers and examinees operating in the face of real consequences.

Some past research examining modality effects on OPI scores has required testers to interview the same examinee in both a face-to-face and telephonic modality. This design feature could prompt testers to make a concerted effort to offer an equivalent interview across the two modality types. Our results suggest that this type of consistency may not exist in the real world, when examinees are not tested twice and interviewers/examinees are not conscious of

their participation in a study.

Theory indicates that the relatively low scores in the telephonic medium may occur in part because rapport deficits cause telephonic testers to feel less connected to the examinee than face-to-face testers. This psychological distance is presumed to lower telephonic ratings by reducing the degree to which telephonic testers give examinees “the benefit of the doubt.” An earlier study investigating modality effects on OPI scores altered the OPI protocol, for research purposes, in a manner that prohibited interviewers from scoring examinees. This protocol modification could have prevented differences between face-to-face and telephonic OPI scores from surfacing because interviewer ratings were removed from the process. By examining the OPI in a naturalistic environment, our research avoided this external validity threat.

#### *Implications for Research and Practice*

The present study has several notable implications for research and practice related to interviews and foreign language assessment. From a practical standpoint, it is presently unwise to assume that a face-to-face and telephonic examinee with the same OPI score operate at equivalent levels of language proficiency. Telephonic examinees appear to be at a disadvantage when compared to others who have been interviewed in person. The ramifications of this point are quite serious. OPIs may be used for a number of important purposes, including the assessment of applicants and the evaluation of training. The failure to recognize the potential effects of modality on interview scores assigned in these and other contexts can lead to faulty decisions, misinformed conclusions, and unfortunate consequences. Because personnel decisions (e.g., graduation from language training in the military) are linked to OPI scores, these unfortunate consequences have real impacts

on the careers and lives of Soldiers, Sailors, Airmen, and Marines.

Discrepancies between face-to-face and telephonic OPIs can also adversely affect examinees’ fairness perceptions. Procedural justice refers to the perceived fairness of the processes organizations use to allocate rewards and determine outcomes (Cohen-Charash & Spector, 2001). As such, reactions to tests administered by organizations are an important determinant of procedural justice. A number of studies have demonstrated linkages between justice perceptions and important outcomes such as job performance, counterproductive work behaviors, organizational citizenship behavior, satisfaction, and organizational commitment (e.g., Cohen-Charash & Spector, 2001). The implication is that work outcomes may suffer to the extent that telephonic examinees feel their OPI modality put them at an unfair disadvantage, especially if OPI scores are used for high-stakes decisions or compensation.

This study suggests that either the telephonic medium reduces interview ratings, the face-to-face medium inflates ratings, or both. Research is needed to determine the source of the measurement error in question. It is possible that one OPI modality induces systematic error variance into the assessment of foreign language speaking skills. To the extent that such error is understood and identified, scores within one modality could potentially be adjusted to equate them to scores within the alternate testing modality. Another possibility is that telephonic and face-to-face OPIs measure different constructs, rendering any comparisons of scores across the two modalities faulty and potentially unfair. In this case, any decision rules applied uniformly to scores within either or both modalities would be faulty and potentially unfair as well.

Research examining when and why

modality affects interview scores would also be informative. Several potential reasons for the modality effect exist. One possibility has to do with the influences of interview modality on communication processes (e.g., perhaps sound quality issues and a lack of visual cues impede understanding during telephonic interviews). The interview content may also be a factor, with interviewers asking different questions in one format versus the other. Moreover, examinees may perform differently in one format versus the other due to variation in their confidence levels and the rapport they develop with the interviewer. Finally, various factors that distinguish the two modalities (e.g., the psychological distance separating the tester from the examinee) may alter interviewers' implicit scoring policies. Clearly, research is needed to determine which if any of these factors drive modality effects on interview ratings. Moreover, studies examining whether situational determinants and other variables (e.g., interviewer experience) moderate the effect of modality on interview scores would be instructive. Ultimately, a better understanding of these issues can inform strategies for minimizing the disparity between face-to-face and telephonic OPI scores.

In reviewing the results of this study, it is important to avoid assuming that telephonic scores are less valid just because they appear to be lower. While the present study suggests that lower scores stem from telephonic versus face-to-face interviews, it provides no insights on the relative accuracy of judgments resulting from the two interview formats. Do telephonic interviews underestimate proficiency or do face-to-face interviews overestimate it? Perhaps both problems occur. Additional research is needed to clarify this matter.

With regard to the accuracy issue, it is interesting to consider whether the

telephone's exclusion of visual cues may in fact be a desirable interview feature. McKay, Curtis, Snyder, and Satterwhite (2005) maintain that raters' encoding of an interviewee's nonverbal behavior is a potential source of measurement error which should be extracted from the rating process due to its influence on interview evaluations. Others have suggested that the telephone could increase accuracy by filtering out cues such as race and attractiveness, which can inappropriately bias interviewers' evaluations (Anderson, 2003; Silvester et al., 2000; Straus et al., 2001). In the absence of cues that signal social and contextual information, interviewers may be forced to evaluate examinees' speech samples more objectively. This assertion is supported by anecdotal evidence from Jackson's (1999) OPI research: "A few of the Russian testers ... noted that if they could not see the examinee, they could not be influenced by his/her appearance. Another Russian tester wrote that she would rather rate audiotapes than videotapes because 'the candidate cannot charm me and I will be more objective.'" (Jackson, 1999, p. 21).

The ability to receive nonverbal cues and gestures can fill gaps in verbal communication when examinees are interpreting testers' speech and vice versa. This also has implications when researching the relative accuracy of telephonic and face-to-face interviews. The lack of visual cues during a telephonic interview could be a problem if face-to-face fluency is the end goal, or it could be a positive assessment feature if telephonic fluency is most desirable. In other words, the influence of nonverbal cues on interviewer judgments may or may not be a biasing factor, depending on the type of oral proficiency that needs to be assessed. Call center operators communicating in a foreign language must be orally proficient in the absence of visual cues and gestures.

Conversely, military personnel who are able to appropriately convey and interpret nonverbal, expressive complements to oral communications will probably be more successful in the field than those who lack this ability. Future research investigating the relative accuracy of telephonic and face-to-face OPI assessments should begin with a clear conceptualization of the oral proficiency type (cue deprived versus cue enabled) that is of interest.

### *Conclusion*

This research begins to shed light on a problem of great practical concern. The findings suggest that the telephonic OPI yields lower ratings than its face-to-face analogue. In the near term, great caution should be exercised when comparing or combining assessments generated from these two different interview types. In the future, additional research should be conducted to confirm the modality effect uncovered in this study and pinpoint why it occurs. Such work can help identify the conditions under which OPI scores are likely to be most accurate. Ultimately, it can also inform strategies for equalizing the two interview types.

## References

- Adams, M. (1978). Measuring foreign language speaking proficiency: A study of agreement among raters. In J. L. D. Clark (ed.), *Direct testing of speaking proficiency: Theory and application* (pp. 131–49). Princeton, NJ: Educational Testing Service.
- Anderson, N. (2003). Applicant and recruiter reactions to new technology in selection: A critical review and agenda for future research. *International Journal of Selection and Assessment, 11*, 121-136.
- Bachman, L. F., & Palmer, A. S. (1981). The construct validity of the FSI oral interview. *Language Learning, 31*, 67–86.
- Brown, A. (2003). Interviewer variation and the co-construction of speaking proficiency. *Language Testing, 20*, 1-25.
- Caligiuri, P., & Lazarova, M. (2002). A model for the influence of social interaction and social support on female expatriates' cross-cultural adjustment. *International Journal of Human Resource Management, 13*, 761-772.
- Caligiuri, P., Phillips, J., Lazarova, M., Tarique, I., & Burgi, P. (2001). The theory of met expectations applied to expatriate adjustment: The role of crosscultural training. *International Journal of Human Resource Management, 12*, 357-372.
- Cascio, W. F. (2003). Changes in workers, work, and organizations. In W. C. Borman, D. R. Ilgen, & R. J. Klimoski (Eds.) *Handbook of psychology* (Vol. 12, pp. 401-422). Hoboken, NJ: John Wiley & Sons, Inc.
- Callaghan, G., & Thompson, P. (2002). 'We recruit attitude': The selection and shaping of routine call centre labour. *Journal of Management Studies, 39*, 233-254.
- Chalhoub-Deville, M., & Fulcher, G. (2003). The oral proficiency interview: A research agenda. *Foreign Language Annals, 36*, 498-506.
- Chan, D., Schmitt, N., Jennings, D., & Sheppard, L. (1999). Developing measures of basic job-relevant English proficiency for the prediction of job performance and promotability. *Journal of Business and Psychology, 14*, 305-318.
- Chao, G. T., & Sun, Y. J. (1997). Training needs for expatriate adjustment in the People's Republic of China. In Z. Aycan (Ed.), *New approaches to employee management: Vol. 4 expatriate management: Theory and research* (pp. 207-226). Greenwich, CT: JAI Press.
- Chittum, R. (2004, May 5). Rise in offshore jobs expected. *Wall Street Journal*, p. B6.
- Clark, J. D. L. (1986). *A study of the comparability of speaking proficiency interview ratings across three government language training agencies*. Washington, DC: Center for Applied Linguistics.
- Cohen-Charash, Y., & Spector, P. E. (2001). The role of justice in organizations: A meta-analysis. *Organizational Behavior and Human Decision Processes, 86*, 278-321.

- Dandonoli, P., & Henning, G. (1990). An investigation of the construct validity of the ACTFL Oral Proficiency Guidelines and Oral Interview Procedure. *Foreign Language Annals*, 23, 11–22.
- Drummond, K., & Hopper, R. (1991). Misunderstanding and its remedies: Telephone miscommunication. In H. Giles and N. Coupland (Eds.), *Miscommunication and problematic talk* (pp. 301-314). Thousand Oaks, CA: Sage.
- Flynn, L. J. (2003, December 8). Companies sending work aboard are learning cultural sensitivity—to their American customers. *The New York Times*, p. C4.
- Glisan, E. W., & Foltz, D. A. (1998). Assessing students' oral proficiency in an outcome-based curriculum: Student performance and teacher intuitions. *The Modern Language Journal*, 82, 1-18.
- Green, R., & Wall, D. (2005). Language testing in the military: Problems, politics, and progress. *Language Testing*, 22, 379-398.
- Hammers, M. (2005). Wyndham looks to leap language gap. *Workforce Management*, p. 17
- Halleck, G. B. (1995). Assessing oral proficiency: A comparison of holistic and objective measures. *The Modern Language Journal*, 79, 223-234.
- Herzog, M. (2003). Impact of the proficiency scale and the oral proficiency interview on the foreign language program at the Defense Language Institute Foreign Language Center. *Foreign Language Annals*, 36, 566-571.
- Jackson, G. L. (1999, February). *Oral proficiency testing modality study*. [Technical Report 99-01]. Presidio of Monterey, CA: Defense Language Institute Foreign Language Center Research and Analysis Division.
- Johnson, M. (2001). *The art of non-conversation: A reexamination of the validity of the oral proficiency interview*. New Haven, CT: Yale University Press.
- Magnan, S. S. (1987). Rater reliability of the ACTFL Oral Proficiency Interview. *The Canadian Modern Language Review*, 43, 267–76.
- Malone, M. E. (2003). Research on the oral proficiency interview: Analysis, synthesis, and future directions. *Foreign Language Annals*, 36, 491-497.
- McKay, P. F., Curtis, J. R., Jr., Snyder, D., & Satterwhite, R. (2005, April). *Panel ratings of tape-recorded interview responses: Interrater reliability? Racial differences?* Paper presented at the 20<sup>th</sup> annual conference of the Society for Industrial and Organizational Psychology, Los Angeles, California.
- McNamara, T. (1998). Policy and social considerations in language assessment. *Annual Review of Applied Linguistics*, 18, 304-319.
- Noe, N. A. (2005). *Employee Training & Development* (3<sup>rd</sup> ed.), Burr Ridge, IL: Irwin McGraw-Hill.
- O'Sullivan, B. (2002). Learner acquaintanceship and oral proficiency test pair-task performance. *Language Testing*, 19, 277-295.

- Orme, D. R., Brehm, W., & Ree, M. J. (2001). Armed forces qualification test as a measure of premorbid intelligence. *Military Psychology, 13*, 187-197.
- Paronto, M. E., Bauer, T. N., & Truxillo, D. M. (2003, April). *Applicant reactions to three screening methods: What do candidates prefer?* Paper presented at the 18<sup>th</sup> annual conference of the Society for Industrial and Organizational Psychology, Orlando, Florida.
- Pristin, T. (2003, October 8). One victim when jobs go overseas: U.S. office space. *The New York Times*, p. C8.
- Rovira, J. D. (2003). *Importance of foreign language capabilities*. (ALMAR 072/03). Washington, DC: Author. Retrieved July 10, 2006, from <http://www.marines.mil/almars/almar2000.nsf/1babcf316f87f38c852569b8008017e7/952d4b8516b446a585256df8006a75f9?OpenDocument>
- Rutter, D. R., & Stephenson, G. M. (1977). The role of visual communication in synchronizing conversation. *European Journal of Social Psychology, 7*, 29-37.
- Silva, J. M., & White, L. A. (1993). Relation of cognitive aptitudes to success in foreign language training. *Military Psychology, 5*, 79-93.
- Silvester, J., & Anderson, N. R. (2003). Technology and discourse: A comparison of face-to-face and telephone employment interviews. *International Journal of Selection and Assessment, 11*, 206-214.
- Silvester, J., Anderson, N., Haddleton, E., Cunningham-Snell, N., & Gibb, A. (2000). A cross-modal comparison of telephone and face-to-face selection interviews in graduate recruitment. *International Journal of Selection and Assessment, 8*, 16-21.
- Smith, R. A., & Frangos, A. (2004, June 2). Outsourcing likely to slow office rebound; trend undercuts demand for space, research finds; suburbs lose call centers. *The Wall Street Journal*, p. B4.
- Stansfield, C. W., & Kenyon, D. M. (1992). The development and validation of a simulated oral proficiency interview. *The Modern Language Journal, 76*, 129-141.
- Straus, S. G., Miles, J. A., & Levesque, L. L. (2001). The effects of videoconference, telephone, and face-to-face media on interviewer and applicant judgments in employment interviews. *Journal of Management, 27*, 363-381.
- Surface, E. A., & Dierdorff, E. C. (2003). Reliability and the ACTFL oral proficiency interview: Reporting indices of interrater consistency and agreement for 19 languages. *Foreign Language Annals, 36*, 507-519.
- Surface, E.A., Dierdorff, E. C., & Donnelly, J. (2004, April). *Modeling second language proficiency change for US Special Operations personnel*. Poster presented at the 19th annual meeting of the Society of Industrial and Organizational Psychology, Chicago, IL.

- Surface, E. A., Poncheri, R. M., & Bhavsar, K. S. (2006, March). *Investigating the reliability and validity of the English ACTFL OPIc<sup>®</sup> with a sample of Korean employees: The ACTFL OPIc<sup>®</sup> validation study technical report*. Raleigh, NC: Surface, Ward, & Associates.
- Surface, E. A., Poncheri, R. M., Lemmond, G., Shetye, T. (2005). *SOF language transformation strategy needs assessment project: Final project report* (Technical Report #20040606). Raleigh, NC: Surface, Ward, & Associates.
- Swender, E. (2003). Oral proficiency testing in the real world: Answers to frequently asked questions. *Foreign Language Annals*, 36, 520-526.
- Takeuchi, R., Yun, S., & Russell, J. (2002). Antecedents and consequences of the perceived adjustment of Japanese expatriates in the USA. *International Journal of Human Resource Management*, 13, 1224-1244.
- Thompson, I. (1995). A study of interrater reliability of the ACTFL Oral Proficiency Interview in five European languages: Data from ESL, French, German, and Spanish. *Foreign Language Annals*, 28, 407-422.
- United States Department of Defense (2005, February 1). A call to action for national foreign language capabilities. *Proceedings of the National Language Conference*, Retrieved July 10, 2006, from [http://www.nlconference.org/docs/White\\_Paper.pdf](http://www.nlconference.org/docs/White_Paper.pdf)
- Vandergrift, L. (1997). The Cinderella of communication strategies: Reception strategies in interactive listening. *The Modern Language Journal*, 81, 494-505.
- Vina, G., & Mudd, T. (2003, November 5). Call centers migrate to India, and north of England loses jobs. *Wall Street Journal*, p. 1.

*Table 1.* Defense Language Institute: Correlations Among Study Variables

Variable	<i>M</i>	<i>SD</i>	<i>N</i> <sup>1</sup>	1	2	3	4	5	6
1. Interview Modality <sup>2</sup>	n/a	n/a	32	–					
2. Test Year	n/a	n/a	32	.29	–				
3. Language Difficulty	2.81	1.06	32	-.60**	-.17	–			
4. Cognitive Ability	81.70	14.22	30	-.25	-.03	.29	–		
5. Language Aptitude	95.74	22.35	19	.13	.31	-.02	.09	–	
6. Interview Score (OPI)	2.84	1.42	32	-.43*	-.46**	.15	.24	-.15	–

<sup>1</sup> Sample size differed across variables due to incomplete cases.

<sup>2</sup> Face-to-face interviews were coded as “1,” and telephonic interviews were coded as “2.”

\* $p < .05$     \*\* $p < .01$

*Table 2.* Defense Language Institute: ANCOVA Analysis Examining the Influence of Modality, Holding Other Relevant Variables Constant

Predictor	Covariates	<i>F</i>	<i>p</i>	Effect Size ( $\eta^2_p$ )
	Test Year	.03	.863	.003
	Language Difficulty	.03	.869	.003
	Cognitive Ability	.04	.846	.004
	Language Aptitude	1.19	.298	.098
Interview Modality		18.48	.001	.627

*Note.* Sample size was reduced due to incomplete cases. This analysis ( $N = 17$ ) included 6 face-to-face and 11 telephonic examinees.

Table 3. USAJFKSWCS: Correlations Among Study Variables

Variable	<i>M</i>	<i>SD</i>	<i>N</i> <sup>1</sup>	1	2	3	4	5	6
1. Interview Modality <sup>2</sup>	n/a	n/a	52	–					
2. Test Year	n/a	n/a	49	.15	–				
3. Language Difficulty	2.85	1.14	52	.00	-.15	–			
4. Cognitive Ability	79.45	15.01	49	.08	.09	.12	–		
5. Language Aptitude	100.83	19.91	36	-.19	.21	.22	.58**	–	
6. Interview Score (OPI)	2.52	1.29	52	-.50**	.01	-.32*	-.03	.28	–

<sup>1</sup> Sample size differed across variables due to incomplete cases.

<sup>2</sup> Face-to-face interviews were coded as “1,” and telephonic interviews were coded as “2.”

\* $p < .05$     \*\* $p < .01$

*Table 4.* USAJFKSWCS: Matched Sample Comparison Between Face-to-Face and Telephonic Examinees

	Face-to-Face Examinees ( <i>N</i> = 26)		Telephonic Examinees ( <i>N</i> = 26)		<i>t</i>	<i>df</i> <sup>1</sup>	<i>p</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>			
Test Year	1999.70	2.12	1999.83	1.30	-0.30	22	.766
Language Difficulty <sup>2</sup>	2.85	1.16	2.85	1.16	n/a	n/a	n/a
Cognitive Ability	78.13	14.75	78.35	14.97	-0.87	22	.396
Language Aptitude	106.08	20.99	106.25	20.81	-0.23	11	.825
Interview Score (OPI) <sup>3</sup>	3.15	1.12	1.88	1.14	5.31	25	<.001

<sup>1</sup> Sample size differed across variables due to incomplete cases.

<sup>2</sup> A t-test was not computed for language difficulty because every telephonic case matched its face-to-face counterpart exactly, producing identical face-to-face and telephonic group means.

<sup>3</sup> The median face-to-face OPI score was 3.00 and the median telephonic score was 2.00.

*Table 5.* ANCOVA Analysis Examining the Influence of Modality, Holding Other Relevant Variables Constant

Predictor	Covariates	<i>F</i>	<i>p</i>	Effect Size ( $\eta^2_p$ )
	Test Year	1.19	.287	.047
	Language Difficulty	36.88	<.001	.606
	Cognitive Ability	0.67	.421	.027
	Language Aptitude	16.25	<.001	.404
Interview Modality		44.56	<.001	.650

*Notes.* Sample size was reduced due to incomplete cases. This analysis ( $N = 30$ ) included 9 face-to-face and 21 telephonic examinees.

## **ABOUT SWA CONSULTING INC.**

SWA Consulting Inc. (formerly Surface, Ward, and Associates) provides analytics and evidence-based solutions for clients using the principles and methods of industrial/organizational (I/O) psychology. Since 1997, SWA has advised and assisted corporate, non-profit and governmental clients on:

- Training and development
- Performance measurement and management
- Organizational effectiveness
- Test development and validation
- Program/training evaluation
- Work/job analysis
- Needs assessment
- Selection system design
- Study and analysis related to human capital issues
- Metric development and data collection
- Advanced data analysis

One specific practice area is analytics, research, and consulting on foreign language and culture in work contexts. In this area, SWA has conducted numerous projects, including language assessment validation and psychometric research; evaluations of language training, training tools, and job aids; language and culture focused needs assessments and job analysis; and advanced analysis of language research data.

Based in Raleigh, NC, and led by Drs. Eric A. Surface and Stephen J. Ward, SWA now employs close to twenty I/O professionals at the masters and PhD levels. SWA professionals are committed to providing clients the best data and analysis with which to make solid data-driven decisions. Taking a scientist-practitioner perspective, SWA professionals conduct model-based, evidence-driven research and consulting to provide the best answers and solutions to enhance our clients' mission and business objectives. SWA has competencies in measurement, data collection, analytics, data modeling, systematic reviews, validation, and evaluation.

For more information about SWA, our projects, and our capabilities, please visit our website ([www.swa-consulting.com](http://www.swa-consulting.com)) or contact Dr. Eric A. Surface ([esurface@swa-consulting.com](mailto:esurface@swa-consulting.com)) or Dr. Stephen J. Ward ([sward@swa-consulting.com](mailto:sward@swa-consulting.com)).