Research Report

We Are What We Repeatedly Do—But in What Context? The Role of Situational Factors in Assessing Personality via Nonverbal Behavior

Jennifer Klafehn

Harrison Kell

Jessica Andrews

Patrick Barnwell

Saad Khan

July 2016

ETS Research Report Series

Since its 1947 founding, ETS has conducted and disseminated scientific research to support its products and services, and to advance the measurement and education fields. In keeping with these goals, ETS is committed to making its research freely available to the professional community and to the general public. Published accounts of ETS research, including papers in the ETS Research Report series, undergo a formal peer-review process by ETS staff to ensure that they meet established scientific and professional standards. All such ETS-conducted peer reviews are in addition to any reviews that outside organizations may provide as part of their own publication processes. Peer review notwithstanding, the positions expressed in the ETS Research Report series and other published accounts of ETS research are those of the authors and not necessarily those of the Officers and Trustees of Educational Testing Service.

The Daniel Eignor Editorship is named in honor of Dr. Daniel R. Eignor, who from 2001 until 2011 served the Research and Development division as Editor for the ETS Research Report series. The Eignor Editorship has been created to recognize the pivotal leadership role that Dr. Eignor played in the research publication process at ETS.

RESEARCH REPORT

We Are What We Repeatedly Do—But in What Context? The Role of Situational Factors in Assessing Personality via Nonverbal Behavior

Jennifer Klafehn, Harrison Kell, Jessica Andrews, Patrick Barnwell, & Saad Khan

Educational Testing Service, Princeton, NJ

While the vast majority of assessments designed to measure noncognitive skills rely heavily on self-report, concerns surrounding faking and response bias have led to an increased demand for the development of new and innovative methods by which to measure such traits, particularly in high stakes contexts. The focus of this report is to address one such method—the inference of noncognitive skills, namely personality, through nonverbal behavior—and the role contextual factors may play in influencing the validity of conclusions drawn from its application. Specifically, this report discusses the various ways in which characteristics of the task (e.g., level of cooperation, physical effort) may facilitate or hinder the assessment of personality when personality is being inferred through observable behavior. This discussion is dovetailed by a review of research from the nonverbal assessment and interpersonal task literatures, as well as a synthesis of these literatures to highlight the degree to which situational factors (as manifested through different tasks) influence the assessment of personality via nonverbal behavior. Additionally, the potential value of using noninvasive tools, such as the sociometer, to collect nonverbal behavioral data during these tasks is discussed. The report concludes with a brief commentary on the feasibility of using nonverbal methods to assess noncognitive skills, with a specific focus on the extent to which structuring the context to elicit certain behaviors may influence the validity and robustness of nonverbal responses as a measurement source.

Keywords Nonverbal behavior; personality; assessment; task; sociometer

doi:10.1002/ets2.12114

The past 25 years of research have demonstrated the importance of noncognitive skills[1] for predicting many important life outcomes (Friedman & Kern, 2014; Heckman, Humphries, & Kautz, 2014; Roberts, Kuncel, Shiner, Caspi, & Goldberg, 2007). Personality traits are a major constituent of the noncognitive skill domain and are associated with a wide variety of workplace criteria, in particular, including job performance (Barrick, Mount, & Judge, 2001; Dudley, Orvis, Lebiecki, & Cortina, 2006; Hogan & Holland, 2003), leadership effectiveness (Judge, Bono, Ilies, & Gerhardt, 2002), job satisfaction (Judge, Heller, & Mount, 2002), contextual performance (Borman, Penner, Allen, & Motowidlo, 2001; Organ & Ryan, 1995), motivation (Judge & Ilies, 2002), and counterproductive work behavior (Berry, Ones, & Sackett, 2007). As a result of these and other related findings, organizations are becoming increasingly interested in using personality as a criterion upon which to base their selection of future employees. While the majority of assessments purporting to measure personality traits rely heavily on self-report, concerns surrounding faking and response bias have led to an increased demand for the development of new and innovative methods by which to measure such traits, particularly in high-stakes contexts (Stark, Chernyshenko, Chan, Lee, & Dragow, 2001). The focus of this report is to address one such method—the inference of personality through nonverbal behavior—and the role contextual factors may play in influencing the validity of conclusions drawn from its application.

The notion that one's personality manifests itself, in part, through one's behavior is not a new concept. Many of the ways in which individuals describe themselves or others often reference attributes that are behavioral in nature. For example, people who are conscientious are described as keeping an organized living space and arriving on time to scheduled events, whereas people who are extraverted are described as talking a lot and spending time at social gatherings. Beyond their purely descriptive value, behaviors have also served as the foundational element upon which a number of important psychological theories have been based. For instance, in developmental psychology, the theories of temperament (Thomas & Chess, 1977) and attachment style (Bowlby, 1969, 1973, 1979) are almost wholly dependent on observable behavioral

Corresponding author: Jennifer Klafehn, E-mail: jklafehn@ets.org

differences in children (e.g., children who are fussy or easily upset are described as having a difficult temperament; children who demonstrate anxiety when their mother leaves and anger when she returns are described as having an anxious ambivalent attachment style). Likewise, many definitions of personality make direct reference to behavior as the distinctive, outward expression of one's traits (see Allport, 1961; Funder, 2006). Despite widespread acknowledgment of the relationship between personality and behavior, however, the assessment of personality has been dominated by measures that ask individuals to reflect and report on (as opposed to demonstrate) their actions. This trend is largely due to the assumption that individuals "know" their personality better than anyone else—after all, individuals are privy to the thoughts and feelings that coincide with and serve to motivate their decisions and subsequent behavior (Markus, 1983). As previously noted, however, personal bias and interpretation often influence how individuals will respond when asked questions about themselves, a concern that could be, in many ways, mitigated if such assessments were completed by another party, such as one's peers or unbiased raters.

There has been substantial research demonstrating the utility of peer- or rater-based assessments of individuals' personality, especially in contexts where bias or faking may be present (see Connolly, Kavanagh, & Viswesvaran, 2007). However, due to the logistical considerations that coincide with administering assessments to individuals other than the target, as well as the lack of standardization regarding whom targets consider their peers, most peer-based assessments are not practical to implement in organizational settings. On the other hand, the measurement or inference of personality via nonverbal behavior not only can be more practical to implement than peer ratings (e.g., consider the use of noninvasive technology, such as sociometers), but is also capable of circumventing many of the issues associated with the use of self-report methods. First, the performance of behavior is not easily faked. Unlike self-report, which, when faked, only requires the test taker to consider what the optimal response to an item should be and then answer accordingly (Holtgraves, 2004; Sudman, Bradburn, & Schwarz, 1996), behavioral responses are often automatic and can occur outside one's awareness. Even in cases where individuals take direct measures to correct or control their behavior, very few are capable of completely masking or overriding the more nuanced reactions that occur in response to environmental stimuli. Thus, the ease with which test takers could fake those behaviors identified as desirable presumably would be lower than their ability to fake responses on a self-report inventory. A second advantage to using nonverbal measures of personality is that they do not require the comprehension and interpretation of written test items. When administering standard paper-and-pencil inventories, it is assumed that test takers are capable of not only reading and understanding test items, but also correctly interpreting what is being referenced in those items. While issues coinciding with the latter assumption may be partially mitigated through the use of anchoring vignettes (King, Murray, Salomon, & Tandon, 2004), the former assumption necessarily excludes populations of individuals from whom the collection of test data may be desired (e.g., children, illiterate adults, non-English speakers; Paunonen & Ashton, 1998). There is also concern that the content of test items too often reflects a Western, White, middle-class bias (Helms, 1992; Lonner, 1981; Solano-Flores & Nelson-Barber, 2001). Thus, for those individuals whose cultural experiences are neither Western nor White nor middle class in origin, performance on an assessment that includes culturally biased items may differ systematically from the performance of their Western, White, middle-class counterparts (see Greenfield, 1997). Although differences likely exist in the way personality is expressed nonverbally across different cultures (Paunonen & Ashton, 1998), as well as the context or scenario in which nonverbal behavior is observed (a concern addressed later in this report), the method of inferring personality through observable behavior does not necessitate that test takers demonstrate their comprehension of or familiarity with test item content in order to be effectively assessed. The third and final advantage to using nonverbal measures of personality is that they can be administered noninvasively and often concurrently with other assessments. For example, in prior studies, researchers have employed various tools such as video recordings (DeGroot & Gooty, 2009; DeGroot & Motowidlo, 1999; Motowidlo & Burnett, 1995), audio recordings (Motowidlo & Burnett, 1995), email and text message content (Chittaranjan, Blom, & Gatica-Perez, 2011), eye-tracking devices (Rauthmann, Seubert, Sachse, & Furtner, 2012), Bluetooth technology (Chittaranjan et al., 2011), automated feature extraction (Naim, Tanveer, Gildea, & Hoque, 2015), and sociometers (e.g., Sociometric Badges; Olguín-Olguín, 2007) to covertly capture nonverbal elements of participants' behavior while they were actively engaged in other activities. Some of these tools, such as the automated feature extraction technology and sociometers, can be implemented without necessitating that a third-party judge or rater assess and evaluate nonverbal data, further increasing their practical value within organizational settings. Not surprisingly, organizations have already begun to implement some of these tools, the sociometer, in particular, as part of their employee performance evaluation system (Waber, Olguín-Olguín, Kim, & Pentland, 2008). In this sense, nonverbal measures constitute

a practical addition to nearly any assessment context, as they do not detract from nor compromise test takers' ability to complete an assessment but still provide administrators with additional information that may be useful in determining test takers' noncognitive skills.

Whereas the number of studies demonstrating the validity of inferring personality through nonverbal behavior is voluminous, little to no research has explored the role of situational factors on the measurement of these behaviors and the effects such measurement may have on the subsequent inference of noncognitive skills. This lack of research is particularly surprising given the decades-long debate over the extent to which behavior is determined by personality as opposed to the situation or context (Mischel, 1968; see also Kenrick & Funder, 1988). Though it is not the intent of this paper to comment on or provide justification for or against one side of this debate versus another, it is the nature of the debate itself that brings to light an important question relevant to the measurement of noncognitive skills via nonverbal methods. Specifically, in what ways does context influence the assessment of noncognitive skills when those skills are being inferred through observable behavior? Several studies have shown that responses on self-reported assessments of personality are highly sensitive to changes in context, one of the most frequently demonstrated examples of which is the shift in responding when participants are being assessed in high- versus low-stakes contexts (e.g., Ellingson, Sackett, & Connelly, 2007). While faking or social desirability may not be as great a concern when measuring personality via nonverbal behavior, certain situations may exist that elicit more valid or robust behavioral responses than others. Furthermore, if interest in assessing noncognitive skills via nonverbal behavior continues to grow, it will be essential to determine whether variance in the "observability" of behavior exists across situations, and, if it does, which situations are more or less likely to provide better measurement opportunities for particular noncognitive skills. The goal of this paper is to present a theoretical premise upon which future discussions and empirical investigations of the relationship between context and nonverbal behavior can be based. Specifically, this paper features a review of research from the nonverbal assessment and interpersonal task literatures, as well as a synthesis of these literatures to highlight the degree to which situational factors (as manifested through different tasks) may play a role in the assessment of noncognitive skills via nonverbal behavior. Included in this synthesis is a brief mention of some recent innovations in the development of tools, the Sociometric Badge, in particular, that are explicitly designed to capture nonverbal behavior in a noninvasive yet maximally informative way. This paper concludes with a discussion of the feasibility of using nonverbal methods to assess noncognitive skills, with a specific focus on the extent to which structuring the context to elicit certain behaviors may influence the validity and robustness of nonverbal responses as a measurement source.

Inferring Noncognitive Skills Through Nonverbal Behavior: A Review

Inferences about psychological characteristics based on people's nonverbal attributes long predates scientific psychology. The humoralist theories of temperament set forth by ancient physicians Hippocrates and Galen specified different types of individuals that were even sometimes depicted in different ways physically, as when presented in artistic works (e.g., the choleric type appearing angry, the melancholic as sad or serious; Dumont, 2010). Phrenology was largely based on making attributions about people's psychological attributes based on the shapes of different parts of their heads (Dumont, 2010). More formal studies of the association between psychological characteristics and nonverbal attributes arose as scientific psychology matured in the 20th century (e.g., Allport & Vernon, 1933; Cleeton & Knight, 1924; Kretschmer, 1925; Paterson, 1930).

Major Concepts

The research literature on nonverbal attributions about personality is large and long-standing, and this review does not attempt to summarize all of it. Instead, major concepts are introduced and defined, and major research findings and designs are discussed, as are more technical issues worth considering that are not as frequently addressed in extant research. Further, this review focuses on topics and findings that are mostly germane to the *zero-acquaintance paradigm* of nonverbal personality assessment, which entails strangers providing judgments of individuals' personality traits, often in laboratory settings (Albright, Kenny, & Malloy, 1988). We focus on this area because it is most relevant to research within applied domains (e.g., predicting job or academic performance).

Studies focusing on nonverbal inferences of personality are often concerned with three major concepts: consensus, self-other agreement, and accuracy (Funder & West, 1993; Leising & Borkenau, 2011). *Consensus* is the extent to which

observers (those who are making the personality judgments) agree in their assessments about targets' (those whose personalities are being judged) standings on the constructs of interest. *Self-other agreement* is the extent to which observers' evaluations agree with targets' self-evaluations. *Accuracy* is the extent to which nonverbal personality evaluations are correct or "true." This review addresses the first two concepts only, as the third quickly ventures into philosophical territory about how to determine "objectively" what someone's "true" personality consists of and what its score is. Relevant to the scope of this review, it is important to note that both consensus and self-other agreement increase as a function of the degree of familiarity between the target and the observer(s) (Connelly & Ones, 2010).

Like self-report measures, nonverbal assessments of personality are concerned with differences in rank-order standings on the traits of interest as opposed to absolute scores. Thus, when there is a high degree of observer consensus regarding judgments of targets' personality traits, this signifies that observers are evaluating and, in turn, rank-ordering the targets in similar ways, rather than observers agreeing on the actual scores assigned to the targets' personalities. This approach is exemplified by the use of various correlational techniques to index consensus and self-other agreement (e.g., Cohen's kappa, intraclass correlation coefficient, Pearson correlation; Leising & Borkenau, 2011). For example, when two observers evaluate the degree of extraversion of each member of a group, they assign each person some score on the trait. Next, agreement can be indexed by computing a correlation coefficient for two evaluators' ratings of extraversion. In the process of computing this correlation coefficient, the targets' raw extraversion scores are transformed into z-scores (Rodgers & Nicewander, 1988), which indicate individuals' rank-order standings among the entire group of extraversion scores. Even if individuals originally receive different absolute scores on extraversion from the two judges, if they are consistently rank-ordered similarly (i.e., tend to be considered *more* or *less* extraverted than others rated, as indicated by z-scores) the correlation coefficient will be relatively high. Indeed, the idea that traits distinguish people from each other is often considered central to their definition (Roberts & Mroczek, 2008).

Two important influences on consensus and self-other agreement are trait evaluativeness and trait observability (Funder, 1995; John & Robins, 1993). *Evaluativeness* refers to how socially desirable a trait is (e.g., friendliness is generally considered to be more socially desirable than hostility) and tends to decrease consensus and self-other agreement (Funder & Colvin, 1988; Funder & Dobroth, 1987; John & Robins, 1993). *Observability* refers to the extent to which the major characteristics of a trait are available to observers in that they are more visible or more frequently occur (Funder, 1995); interrater agreement when judging extraversion is typically high, for example, because a major component of its definition is highly visible: social interactions with others. One content analysis has revealed that personality inventories usually conceptualize traits in terms of the classic tripartite model of cognition, affect, and behavior (Hilgard, 1980; Zillig, Hemenover, & Dienstbier, 2002).[2] Traits differ in the extent to which they can be characterized by these components, with some traits being more amenable to explicit, overt behavioral expressions (e.g., extraversion) as opposed to others that are more amenable to internal mental and emotional states (e.g., neuroticism). Not surprisingly, less observable traits demonstrate lower self-other agreement and consensus than highly observable traits (Connelly & Ones, 2010).

Predominant Research Designs and Findings

Research Designs: Laboratory Settings

By their very nature, evaluations of personality traits via zero-acquaintance designs tend not to be overly complex. In laboratory settings, targets usually perform some standardized task(s), which may or may not involve interacting with a confederate (Borkenau & Liebler, 1992, 1993; Borkenau, Mauer, Riemann, Spinath, & Angleitner, 2004; Borkenau, Riemann, Spinath, & Angleitner, 2006), and then complete a self-report personality measure. Observers then view still images or videotapes of the targets and rate their personalities, often using adjective scales (Goldberg, 1992; Norman, 1963). Less commonly, the observers actually interact with the targets for some period of time before providing their ratings (e.g., Passini & Norman, 1966). More complex designs entail different groups of raters providing personality assessments based on separate aspects of the stimulus materials. For example, while one group of raters may evaluate targets' traits based solely on a videotape with the sound turned off, another group of raters may provide ratings based solely on the audio gleaned from the same videotape, with a third group of raters providing its evaluations based on the entire videotape, and a fourth group providing its ratings based solely on a still image taken from the experimental session. This is essentially a multitrait multimethod (D. T. Campbell & Fiske, 1959) approach to nonverbal personality assessment, with the assumption being that if the ratings of targets' personality traits possess construct validity, they will converge across the various

methodologies. In such designs, particular care must be exercised in creating appropriate still images to avoid sampling behaviors that occur at a low-base rate for the person in question or essentially randomly (e.g., a highly disagreeable person laughs in response to an off-hand comment a research staff member makes). One approach is to include a brief experimental session specifically aimed at obtaining neutral still images, wherein targets only look at the camera and are not asked to perform a task (Borkenau & Liebler, 1993). Additionally, interpretation of convergence (or lack thereof) across methods must be made in light of a trait's degree of observability (e.g., more convergence would be expected for extraversion than neuroticism).

The preceding discussion describes holistic or subjective approaches to personality assessment via nonverbal behavior. A more "objective" approach entails observers providing ratings of targets' specific behaviors and physical attributes, such as attractiveness, body orientation, gait, loudness of voice, and movement speed (Borkenau & Liebler, 1993). Additionally, audio recordings of targets' voices can be analyzed via computer, and information about such characteristics as amplitude variability, pitch, and speech rate can be extracted (DeGroot & Kluemper, 2007; DeGroot & Motowidlo, 1999). Although these objective ratings could be compared directly across observers, a more common approach is for observers also to complete holistic, adjective-based judgments of targets' personalities, allowing for the regression of those subjective ratings onto the objective ratings and the derivation of a model of how those observers are making personality judgments based on those narrow nonverbal cues (i.e., the Brunswik lens model; Brunswik, 1956). An additional approach is for an independent group of observers to provide ratings of the objective nonverbal cues, which are then averaged in an attempt to eliminate measurement error in the form of raters' idiosyncrasies. These "true" measures of the objective cues can then be alternatively used in the regression models.

More recently, researchers have begun to explore other methods by which to capture elements of nonverbal behavior that are both more "objective" and noninvasive than some of the methods just described. One such method focuses on the collection of low-level nonverbal behavioral data from wearable sensors known as sociometers (Choudhury & Pentland, 2003; Olguín-Olguín, 2007; see also Sociometric Badges). Sociometers are small devices, approximately the size of a smartphone, that are worn by participants and are designed to capture patterns in physical activity, speech activity, and proximity, amongst other low-level behavioral cues. Some of the first psychological studies to employ sociometers as a measurement tool focused on modeling the structure and dynamics of social networks (Choudhury & Pentland, 2003; Pentland, 2006). This approach utilizes pattern recognition and machine learning methods, such as support vector machines (Cortes & Vapnik, 1995), to analyze raw sensor data (e.g., body motion, speech, proximity to other devices) in order to detect and make reliable estimates of users' behavior and interactions. As sociometer technology became increasingly sophisticated, researchers began to explore its potential for capturing more nuanced elements of behavior and communication, specifically those that influenced interpersonal interactions, such as mimicry, conversational turn-taking, and activity (Pentland, 2008; Woolley, Chabris, Pentland, Hashmi, & Malone, 2010). For example, one of the newest iterations of the sociometer, the Sociometric Badge (Olguín-Olguín, 2007), has the capacity to collect information on (a) speech features such as volume, tone of voice, and speaking time; (b) body movement features such as energy and consistency; (c) information regarding people nearby who are also wearing a Sociometric Badge; (d) the proximity of Bluetooth-enabled devices; and (e) approximate location information. To date, sociometers have been used in research examining interpersonal behavior and performance across a wide variety of organizational settings, including healthcare (Olguín-Olguín, Gloor, & Pentland, 2009; Olguín-Olguín & Pentland, 2010b), marketing (Waber et al., 2008), sales (Olguín-Olguín & Pentland, 2010a), and information technology (Wu, Waber, Aral, Brynjolfsson, & Pentland, 2008).

Findings

As would be expected per the tenet of observability (see the "Major Concepts" subsection), the greatest consensus among observers' ratings tends to be for extraversion, with consensus among strangers making ratings from viewing videotapes being high: Borkenau and Liebler (1992) reported a Cronbach's alpha of .81 for extraversion ratings, with Borkenau and Liebler (1993) reporting an identical value. Openness to experience and neuroticism, traits with significantly more emphasis on cognition and affect, evince lower consensus, with primary study estimates ranging from .65 to .75 (Borkenau & Liebler, 1992, 1993) and meta-analytic values of .23 (neuroticism) and .30 (openness; Connelly & Ones, 2010). Self-other agreement tends to be consistently lower than consensus among observers, with the uncorrected correlations being, for example, .22 for extraversion and .08 for neuroticism (Connelly & Ones, 2010).

Research Designs: Applied Settings

Many zero-acquaintance studies conducted in laboratory settings feature tasks with presumably limited external validity, such as reading a weather forecast (Borkenau & Liebler, 1992, 1993) or newspaper headlines (Borkenau et al., 2006), and often use undergraduate participants. Designs focused on predicting a practical criterion (e.g., job performance), however, often include tasks with real-world significance (e.g., mock job interviews; Burnett & Motowidlo, 1998), more representative samples (e.g., currently employed managers; Motowidlo & Burnett, 1995), or both, such as recordings of real job interviews upon which actual hiring decisions were based (e.g., Forbes & Jackson, 1980; Gifford, Ng, & Wilkinson, 1985).

Findings

Self-other agreement and observer consensus for personality traits assessed while targets are engaged in some applied task, such as a job interview, tend to be comparable to those demonstrated in more typical laboratory-based research designs (Barrick, Patton, & Haugland, 2000; DeGroot & Gooty, 2009; DeGroot & Kluemper, 2007; Mount, Barrick, & Strauss, 1994). The most notable aspect of observers' ratings of personality traits is their universal superiority to self-ratings for predicting real-world outcomes. For example, a meta-analysis conducted by Oh, Wang, and Mount (2011) showed a mean uncorrected correlation of .15 between self-reported conscientiousness and job performance but a mean uncorrected correlation of .25 for observer ratings of conscientiousness. They further demonstrated that observer ratings contributed substantial incremental validity to predicting job performance beyond self-reports but not vice-versa. A meta-analysis of the same topic (Connelly & Ones, 2010) drew similar conclusions.

Additional Considerations

An enormous amount of primary and meta-analytic evidence suggests that observers' ratings of personality traits are superior to self-reports when predicting real-world criteria. However, this does not mean self-reports should be entirely excluded from the process, as self-report measures can still add a small amount of variance beyond observer ratings (Oh et al., 2011), the influence of which could aggregate over long periods of time or in very large samples (Roberts et al., 2007). Further, it may be worthwhile to take note of individuals whose self-perceptions of personality differ dramatically from how they are perceived externally. First, do these people consistently score lower (or higher?) on interpersonal tasks than those whose self-perceptions are better aligned with those of observers? Second, even if this extreme misalignment has no immediate practical consequence, it may be worthy of study, as the consequences of this discrepancy could have an impact on behavior or maladaptive outcomes in the long term (Pervin, 1994).

The Role of Contextual Features on the Expression of Nonverbal Behavior

Like personality, context, or the set of features associated with a given situation, is an important factor that influences how individuals behave and communicate (e.g., Fischer et al., 2011; Liberman, Samuels, & Ross, 2004; Matsumoto, 2007). It thereby follows that contextual features potentially can influence the degree to which inferences about individuals' personality can be made from behavioral information. Research has shown that interpersonal situations, in particular, exert marked influence on the ways individuals behave (e.g., H. K. Gardner, 2012; W. L. Gardner, Gabriel, & Lee, 1999). For example, some interpersonal contexts may evoke cooperative or agreeable behaviors whereas others elicit competitive or dominant behaviors (Andrews, 2014; Balliet, Li, Macfarlan, & Van Vugt, 2011; Cox, Lobel, & McLeod, 1991). As such, it is difficult to fully understand interaction processes and performance without taking into account the context or nature of the task in which individuals are engaged[3] (Straus, 1999).

Several researchers have developed theoretical frameworks that classify tasks according to a number of contextual features that can affect interaction processes and performance outcomes. These features include, among others, the goal or product associated with the task (e.g., speed, accuracy), relations between the task and the group working on the task (e.g., task difficulty, intrinsic interest), relations among individual group members (e.g., interdependence, competition), the ways in which individual contributions of group members impact the task's outcome or completion of the task (e.g., disjunctive, conjunctive, additive), and the types of behavioral processes involved in completing the task (e.g., discussion, motor coordination, mechanical assembly; Carter, Haythorn, & Howell, 1950; Driskell, Hogan, & Salas, 1987; Laughlin, 1980; Shaw, 1981; Steiner, 1972). For example, Hackman (1968) developed a task classification framework

that distinguished production, discussion, and problem-solving tasks according to the behavioral requirements for each task type. Specifically, production tasks require individuals to work together to generate ideas, discussion tasks require dialogue among group members on a given topic, and problem-solving tasks require individuals to describe and work on carrying out a plan of action. Despite these tasks being classified separately, however, it should be noted they are not necessarily mutually exclusive. In many real-world contexts individuals may engage in more than one type of task over the course of a single session.

One notable classification scheme, McGrath's (1984) task circumplex, integrates aspects of these prior task classification frameworks and has become widely used in group research. McGrath's model of group task types distinguishes four categories of tasks that differ according to the performance processes or behaviors needed to carry out each task: generate, choose, negotiate, and execute. Each of these task categories/processes is divided into two task subtypes. Specifically, the *generate* category includes creativity tasks in which groups generate as many ideas as possible and planning tasks in which groups determine a plan for carrying out a goal. The *choose* category includes intellective tasks that require groups to determine a solution to a problem with a demonstrably correct answer and decision-making tasks that have no demonstrably correct answer, thus requiring groups to reach a preferred solution. The *negotiate* category includes cognitive conflict tasks and mixed-motive tasks in which group members must resolve conflicts of viewpoint or interest. The final category, *execute*, involves contests and performances, each of which involves groups doing manual and psychomotor tasks. These task types are related to each other within a two-dimensional space. The horizontal dimension is associated with the degree to which tasks have conceptual or action performance requirements (McGrath, 1984), whereas the vertical dimension reflects the degree of interdependence required by a task. The vertical dimension is sometimes expressed with a three-level specification of interdependence (i.e., collaboration, coordination, and conflict resolution), with collaboration tasks requiring the least and conflict resolution tasks requiring the most interdependence among group members (Argote & McGrath, 1993).

These dimensions have important implications for the types of nonverbal behaviors one might expect to observe for the various task types. Moving from more conceptual to more action-oriented tasks should conceivably create more opportunities for individuals to display behavioral information. Furthermore, as the degree of interdependence increases, the amount of information richness required for successful completion of the task increases, as well. *Information richness* refers to the amount of supplemental information (e.g., emotional, attitudinal) provided via nonverbal and paraverbal channels beyond explicit denotation of symbols (Hollingshead, McGrath, & O'Connor, 1993). With respect to McGrath's (1984) task circumplex, the eight task types are ordered according to increasing information richness requirements as a result of the degree of interdependence associated with the task type. The tasks in order of increasing information richness requirements (Kerr & Murthy, 2009) are (a) planning tasks (generating plans), (b) creativity tasks (generating ideas), (c) intellective tasks (solving problems with correct answers), (d) decision-making tasks (solving problems without correct answers), (e) cognitive conflict tasks (resolving conflicts of viewpoint), (f) mixed-motive tasks (resolving conflicts of interest), (g) contests/competitive tasks (resolving conflicts of power), and (h) performance/psychomotor tasks (executing performance tasks).

Presumably, tasks associated with higher information richness requirements should be more likely to elicit behavioral information, including nonverbal cues. Take, for example, a creativity task in which individuals are asked to generate a list that includes as many uses as possible for common objects (e.g., coffee cup, wire coat hanger, brick) plus uses other than those for which the object was designed (Harrison, Mohammed, McGrath, Florey, & Vanderstoep, 2003). This sort of task only requires the simple exchange of ideas to generate solutions (i.e., low information richness requirements). Consensus is not necessary so there is no need for individuals to argue their point, and one person can conceivably complete the task with little contribution from others. As a result, evaluative and emotional connotations about messages are not required and may even, in some ways, hinder performance (Hollingshead et al., 1993).

As an alternative example, take a decision-making task in which individuals must determine how to allocate funds to six competing projects that have requested funding from a foundation they, as a group, represent. The competing projects differ in terms of the values to which they appeal. For example, one project with economic appeal proposes to create a tourist bureau to develop advertising and other methods of attracting tourism into the community, whereas another project with religious appeal proposes to establish an additional shelter for the homeless in the community (Watson, DeSanctis, & Poole, 1988). With such a task, there is no demonstrably correct answer, making the ability to reach consensus difficult. There may also be conflicting viewpoints about the best solution given each project's appeal to various values. As a result, group members provide not only factual information, but also messages laden with values, emotions, and

attitudes as they communicate their opinions and rationales for the merit of alternative solutions and also work to reconcile differences in information, attitudes, and opinions (Hollingshead et al., 1993). Such messages should be more likely to elicit nonverbal behaviors such as gestures and changes in paraverbal communication (e.g., volume and tone of voice) than the kind of information exchange necessary for the creativity task described previously (Straus & McGrath, 1994).

These examples illustrate how features associated with different kinds of tasks may influence the extent to which behavioral information is elicited. In particular, tasks that require coordination or resolution of conflicting ideas and for which reaching consensus is difficult create more opportunities for individuals to display behavioral information, including nonverbal and paraverbal cues. In these situations, information richness becomes more important for successful completion of the task. This suggests that if some behaviors are more or less indicative of particular personality traits, it is important to take into account the context or nature of the task in which individuals are being assessed, as some situations may be structured in a way that creates greater opportunity to display certain behaviors.

Assessing Personality via Nonverbal Behavior: Where Do We Go From Here?

The previous sections highlight two important points regarding the measurement of personality via nonverbal behavior: (a) nonverbal behavior can serve as a valid source of information about individuals' personality, and (b) variation in task or performance context may result in variation of the behaviors individuals express and, in turn, the conclusions one is able to draw regarding individuals' personality. Taken together, these points suggest that, while there is merit to using nonverbal measures to assess personality, the administration and application of such measures should be approached in a strategic manner as both could impact the validity of the assessment. Specifically, it is important for assessment administrators, whether they be academic researchers, school counselors, or human resource specialists, to determine not only which traits are of interest or are relevant to measure, but also which context or task is most likely to yield opportunities to measure them.

As noted in this paper's earlier review of the nonverbal personality literature, much of the research that has explored nonverbal measurement of personality has done so within the context of laboratory settings, which tend to employ tasks that are limited in terms of the length of time raters have to observe targets' behavior as well as their practical application to real-world outcomes. While research findings have demonstrated that these laboratory-based tasks tend to be sufficient, insofar as they allow raters to draw, at a minimum, superficial inferences about targets' personality, it is difficult to determine exactly what kinds of inferences raters are making when judging personality traits in situations with applied implications. For example, if raters are viewing videotapes of individuals who are interviewing for a job, it is worthwhile to question whether they are making inferences about targets' overall personalities (i.e., with no context specified), in the immediate context of the job interview itself (i.e., a highly motivated situation), or in the context of performing job activities. People can exhibit enormous intraindividual variation in their traited behaviors (Borkenau et al., 2006; Fleeson, 2001; Fleeson & Gallagher, 2009; Mischel & Shoda, 2008), meaning that inferences made about "personalities in the abstract" may not correspond particularly well to behaviors in specific contexts. Indeed, frame-of-reference self-report personality inventories that specify some context for evaluation (e.g., at work, at home) consistently outperform broadband personality inventories for predicting similarly situated outcomes, such as job performance (Shaffer & Postlethwaite, 2012). Thus, it may be important to design tasks and instructions for raters very carefully when they are evaluating personality traits for applied purposes. One possible option is to design tasks that focus explicitly on the noncognitive skills and abilities identified as critical for the performance of a particular job. For example, emotional stability and agreeableness are often cited as important predictors of performance in jobs that involve extensive interpersonal interaction, such as customer service jobs or jobs involving teamwork (Hough, 1992; Mount, Barrick, & Stewart, 1998). Thus, structuring the measurement task to require maximal information richness and, in turn, more varied and frequent instances of interpersonal behavior will provide raters with more opportunities to observe and judge targets' emotional stability and agreeableness in that particular type of performance context. In addition to specifying the focus of the task, raters also can be directed not to evaluate targets' personality dispositions, per se, but rather to evaluate the targets' behaviors in the given situation only. By providing specific guidance to raters prior to their observation of targets' behavior, administrators can help limit the potential of "surplus inference" being introduced into ratings as much as possible.

At the same time, administrators may be interested in capturing a wider breadth of noncognitive skills. If resources are available, it may be possible to design a series of tasks that purposefully mimic the broad range of situations a performance context may feature (representative design; Albright & Malloy, 2000; Brunswik, 1947), thus allowing for a more direct

assessment of targets' intraindividual behavioral variability. This may be especially important if trait evaluativeness varies greatly across contexts in the criterion domain—that is, the entire criterion domain differs greatly in the desirability of certain trait levels relative to the desirability of the trait in the assessment context or "in general." For example, it is generally socially desirable to be friendly—but it may not be desirable to be friendly if you are a car repossessor—yet would it be desirable to be friendly in an interview to become a car repossessor? In this latter situation there could be a curvilinear effect.

It is equally important to bear in mind, however, that the greater the number of tasks introduced into the measurement context, the more costly and impractical the use of human raters can become. The implementation of noninvasive methods, such as the previously discussed sociometer, thus becomes crucial in these types of measurement designs, as they allow for similar, and, in some cases, more fine-grained data capture but are not subject to the constraints coinciding with the use of human raters. For this reason, sociometers are ideal for any context in which the measurement of nonverbal behavior is desired, as they not only can be implemented at a moment's notice, but are also capable of capturing behavioral nuances (e.g., mirroring, variations in posture) that may otherwise go undetected by human raters.

There is little doubt that noncognitive skills play an important role in predicting performance. Less certain is the ideal method by which to assess those skills that is both valid and resistant to threats related to personal bias, interpretation, and social desirability. Inferring personality through nonverbal behavior may serve as one potential solution to this challenge. As with any type of assessment, however, the context in which the construct of interest is being measured is oftentimes just as important as the measurement of the construct itself. This review illustrates that the inference of personality through nonverbal behavior is no different, and that features of the measurement context may influence not only the degree to which behaviors can be observed, but also the viability of the conclusions drawn from those observations.

Notes

1 We acknowledge that "noncognitive skills" is a problematic term (Duckworth & Yeager, 2015; Kyllonen, Lipnevich, Burrus, & Roberts, 2014). First, it is technically incorrect, as the characteristics often classified as "noncognitive" possess cognitive components (Messick, 1979). Second, it is ambiguous, essentially serving as a catch-all category for seemingly any human attribute not typically thought to be assessed by traditional achievement or "IQ-type" tests, including personality traits, goals, preferences, and motivation (Borghans, Duckworth, Heckman, & ter Weel, 2008; Farkas, 2003; Kautz, Heckman, Diris, ter Weel, & Borghans, 2014). Third, the term "skill" itself is problematic, because it has different definitions in different literatures. For example, J. P. Campbell's (1990, 2012) job performance model differentiates "basic attributes" such as personality traits and cognitive abilities from skills, treating the former as antecedents of the latter, which he defines as currently knowing how (i.e., possessing procedural knowledge) and actually being able to carry out the task in question (Motowidlo & Kell, 2013). Fleishman (1972) also clearly distinguishes skills and basic attributes, describing a skill as the degree of proficiency on a task and an ability as a common process involved in performing multiple tasks. Dawis (1996), on the other hand, does not make such a stark distinction, defining a skill as "an identifiable, repeatable behavior sequence emitted in response to a demand or 'task'" (p. 233) and further stating that "so-called ability tests are actually measures of skills" (p. 234). Wiley (1991) more simply defines skills as "abilities to perform tasks (p. 78). Thus, we acknowledge the many difficulties of the term "noncognitive skills" but believe it is nonetheless valuable in roughly highlighting the distinction between two psychological domains (Messick, 1979): One that has traditionally been invoked in academic achievement settings and one that has not. Further, although we hope that future work will serve to better delimit the domain of so-called noncognitive skills, it is beyond the scope of the current paper to attempt to do so.
2 Wilt and Revelle (2015) identify a fourth component, desire.
3 We acknowledge that task type is not the only contextual feature likely to influence behavior. For example, an entire body of research has been dedicated to the effects of team composition on performance (see Bell, 2007). Given that an in-depth discussion of all contextual influences is beyond the scope of this report, however, we have chosen to focus exclusively on task type due to the lack of research exploring the role of the task in the assessment of personality via nonverbal behavior.

References

Albright, L., Kenny, D. A., & Malloy, T. E. (1988). Consensus in personality judgments at zero acquaintance. *Journal of Personality and Social Psychology*, 55, 387–395.

Albright, L., & Malloy, T. E. (2000). Experimental validity: Brunswik, Campbell, Cronbach, and enduring issues. *Review of General Psychology*, 4, 337–353.

Allport, G. W. (1961). *Pattern and growth in personality*. New York, NY: Holt, Rinehart & Winston.

Allport, G. W., & Vernon, P. E. (1933). *Studies in expressive movement*. New York, NY: Macmillan.

Andrews, J. J. (2014). *An examination of factors influencing erroneous knowledge acquisition in collaborative contexts* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses. (Publication No. 3638115)

Argote, L., & McGrath, J. E. (1993). Group processes in organizations: Continuity and change. In C. L. Cooper & I. T. Robertson (Eds.), *International review of industrial and organizational psychology* (Vol. 8, pp. 333–389). New York, NY: Wiley.

Balliet, D., Li, N. P., Macfarlan, S. J., & Van Vugt, M. (2011). Sex differences in cooperation: A meta-analytic review of social dilemmas. *Psychological Bulletin, 137,* 881–909.

Barrick, M. R., Mount, M. K., & Judge, T. A. (2001). Personality and performance at the beginning of the new millennium: What do we know and where do we go next? *International Journal of Selection and Assessment, 9,* 9–30.

Barrick, M. R., Patton, G. K., & Haugland, S. N. (2000). Accuracy of interviewer judgments of job applicant personality traits. *Personnel Psychology, 53,* 925–951.

Bell, S. T. (2007). Deep-level composition variables as predictors of team performance. *Journal of Applied Psychology, 92,* 595–615.

Berry, C. M., Ones, D. S., & Sackett, P. R. (2007). Interpersonal deviance, organizational deviance, and their common correlates: A review and meta-analysis. *Journal of Applied Psychology, 92,* 410–424.

Borghans, L., Duckworth, A. L., Heckman, J. J., & ter Weel, B. (2008). The economics and psychology of personality traits. *Journal of Human Resources, 43,* 972–1059.

Borkenau, P., & Liebler, A. (1992). Trait inferences: Sources of validity at zero acquaintance. *Journal of Personality and Social Psychology, 62,* 645–657.

Borkenau, P., & Liebler, A. (1993). Convergence of stranger ratings of personality and intelligence with self-ratings, partner ratings, and measured intelligence. *Journal of Personality and Social Psychology, 65,* 546–553.

Borkenau, P., Mauer, N., Riemann, R., Spinath, F. M., & Angleitner, A. (2004). Thin slices of behavior as cues of personality and intelligence. *Journal of Personality and Social Psychology, 86,* 599–614.

Borkenau, P., Riemann, R., Spinath, F. M., & Angleitner, A. (2006). Genetic and environmental influences on person × situation profiles. *Journal of Personality, 74,* 1451–1480.

Borman, W. C., Penner, L. A., Allen, T. D., & Motowidlo, S. J. (2001). Personality predictors of citizenship performance. *International Journal of Selection and Assessment, 9,* 52–69.

Bowlby, J. (1969). *Attachment and loss: Vol. 1. Attachment*. New York, NY: Basic Books.

Bowlby, J. (1973). *Attachment and loss: Vol. 2. Separation: Anxiety and anger*. New York, NY: Basic Books.

Bowlby, J. (1979). *The making and breaking of affectional bonds*. London, England: Tavistock

Brunswik, E. (1947). *Systematic and representative design of psychological experiments*. Berkeley: University of California Press.

Brunswik, E. (1956). *Perception and the representative design of psychological experiments*. Berkeley: University of California Press.

Burnett, J. R., & Motowidlo, S. J. (1998). Relations between different sources of information in the structured selection interview. *Personnel Psychology, 51,* 963–983.

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

Campbell, J. P. (1990). Modeling the performance prediction problem in industrial and organizational psychology. In M. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (Vol. 1, 2nd ed., pp. 687–731). Palo Alto, CA: Consulting Psychologists Press.

Campbell, J. P. (2012). Behavior, performance, and effectiveness in the 21st century. In S. W. J. Kozlowski (Ed.), *The Oxford handbook of organizational psychology* (pp. 159–195). New York, NY: Oxford University Press.

Carter, L., Haythorn, W., & Howell, M. (1950). A further investigation of the criteria of leadership. *Journal of Abnormal and Social Psychology, 45,* 350–358.

Chittaranjan, G., Blom, J., & Gatica-Perez, D. (2011). Mining large-scale smartphone data for personality studies. *Personal and Ubiquitous Computing, 17,* 433–450.

Choudhury, T., & Pentland, A. (2003, October). *Sensing and modeling human networks using the sociometer*. Paper presented at the 7th IEEE International Symposium on Wearable Computers, White Plains, NY.

Cleeton, G. C., & Knight, F. B. (1924). Validity of character judgments based on external criteria. *Journal of Applied Psychology, 8,* 215–231.

Connelly, B. S., & Ones, D. S. (2010). Another perspective on personality: Meta-analytic integration of observers' accuracy and predictive validity. *Psychological Bulletin, 136,* 1092–1122.

Connolly, J. J., Kavanagh, E. J., & Viswesvaran, C. (2007). The convergent validity between self and observer ratings of personality: A meta-analytic review. *International Journal of Selection and Assessment, 15,* 110–117.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning, 20,* 273–297.

Cox, T. H., Lobel, S. A., & McLeod, P. L. (1991). Effects of ethnic group cultural differences on cooperative and competitive behavior on a group task. *Academy of Management Journal, 34,* 827–847.

Dawis, R. V. (1996). Vocational psychology, vocational adjustment, and the workforce: Some familiar and unanticipated consequences. *Psychology, Public Policy, and Law, 2,* 229–248.

DeGroot, T., & Gooty, J. (2009). Can nonverbal cues be used to make meaningful personality attributions in employment interviews? *Journal of Business and Psychology, 24,* 179–192.

DeGroot, T., & Kluemper, D. (2007). Evidence of predictive and incremental validity of personality factors, vocal attractiveness and the situational interview. *International Journal of Selection and Assessment, 15,* 30–39.

DeGroot, T., & Motowidlo, S. J. (1999). Why visual and vocal interview cues can affect interviewers' judgments and predict job performance. *Journal of Applied Psychology, 84,* 986–993.

Driskell, J. E., Hogan, R., & Salas, E. (1987). Personality and group performance. *Review of Personality and Social Psychology: Group Processes and Intergroup Relations, 14,* 91–112.

Duckworth, A. L., & Yeager, D. S. (2015). Measurement matters: Assessing personal qualities other than cognitive ability for educational purposes. *Educational Researcher, 44,* 237–251.

Dudley, N. M., Orvis, K. A., Lebiecki, J. E., & Cortina, J. M. (2006). A meta-analytic investigation of conscientiousness in the prediction of job performance: Examining the intercorrelations and the incremental validity of narrow traits. *Journal of Applied Psychology, 91,* 40–57.

Dumont, F. (2010). *A history of personality psychology: Theory, science, and research from Hellenism to the twenty-first century*. New York, NY: Cambridge University Press.

Ellingson, J. E., Sackett, P. R., & Connelly, B. S. (2007). Personality assessment across selection and development contexts: Insights into response distortion. *Journal of Applied Psychology, 92,* 386–395.

Farkas, G. (2003). Cognitive skills and noncognitive traits and behaviors in stratification processes. *Annual Review of Sociology, 59,* 541–562.

Fischer, P., Krueger, J. I., Greitemeyer, T., Vogrincic, C., Kastenmüller, A., Frey, D., … Kainbacher, M. (2011). The bystander-effect: A meta-analytic review on bystander intervention in dangerous and non-dangerous emergencies. *Psychological Bulletin, 137,* 517.

Fleeson, W. (2001). Toward a structure- and process-integrated view of personality: Traits as density distributions of states. *Journal of Personality and Social Psychology, 80,* 1011–1027.

Fleeson, W., & Gallagher, P. (2009). The implications of Big Five standing for the distribution of trait manifestation in behavior: Fifteen experience-sampling studies and a meta-analysis. *Journal of Personality and Social Psychology, 97,* 1097–1114.

Fleishman, E. A. (1972). On the relation between abilities, learning, and human performance. *American Psychologist, 27,* 1017–1032.

Forbes, R. J., & Jackson, P. R. (1980). Non-verbal behaviour and the outcome of selection interviews. *Journal of Occupational Psychology, 53,* 65–72.

Friedman, H. S., & Kern, M. L. (2014). Personality, well-being, and health. *Annual Review of Psychology, 65,* 719–742.

Funder, D. C. (1995). On the accuracy of personality judgment: A realistic approach. *Psychological Review, 102,* 652–670.

Funder, D. C. (2006). Towards a resolution of the personality triad: Persons, situations, and behaviors. *Journal of Research in Personality, 40,* 21–34.

Funder, D. C., & Colvin, C. R. (1988). Friends and strangers: Acquaintanceship, agreement, and the accuracy of personality judgment. *Journal of Personality and Social Psychology, 55,* 149–158.

Funder, D. C., & Dobroth, K. M. (1987). Differences between traits: Properties associated with interjudge agreement. *Journal of Personality and Social Psychology, 52,* 409–418.

Funder, D. C., & West, S. G. (1993). Consensus, self-other agreement, and accuracy in personality judgment: An introduction. *Journal of Personality, 61,* 457–476.

Gardner, H. K. (2012). Performance pressure as a double-edged sword enhancing team motivation but undermining the use of team knowledge. *Administrative Science Quarterly, 57,* 1–46.

Gardner, W. L., Gabriel, S., & Lee, A. Y. (1999). "I" value freedom, but "we" value relationships: Self-construal priming mirrors cultural differences in judgment. *Psychological Science, 10,* 321–326.

Gifford, R., Ng, C. F., & Wilkinson, M. (1985). Nonverbal cues in the employment interview: Links between applicant qualities and interviewer judgments. *Journal of Applied Psychology, 70,* 729–736.

Goldberg, L. R. (1992). The development of markers for the Big-Five factor structure. *Psychological Assessment, 4,* 26–42.

Greenfield, P. M. (1997). You can't take it with you: Why ability assessments don't cross cultures. *American Psychologist, 52,* 1115–1124.

Hackman, J. R. (1968). Effects of task characteristics on group products. *Journal of Experimental Social Psychology, 4,* 162–187.

Harrison, D. A., Mohammed, S., McGrath, J. E., Florey, A. T., & Vanderstoep, S. W. (2003). Time matters in team performance: Effects of member familiarity, entrainment, and task discontinuity on speed and quality. *Personnel Psychology, 56,* 633–669.

Heckman, J. J., Humphries, J. E., & Kautz, T. (Eds.). (2014). *The myth of achievement tests: The GED and the role of character in American life*. Chicago, IL: University of Chicago Press.

Helms, J. E. (1992). Why is there no study of cultural equivalence in standardized cognitive ability testing? *American Psychologist, 47,* 1083–1101.

Hilgard, E. R. (1980). The trilogy of mind: Cognition, affection, and conation. *Journal of the History of the Behavioral Sciences, 16,* 107–117.

Hogan, J., & Holland, B. (2003). Using theory to evaluate personality and job-performance relations: A socioanalytic perspective. *Journal of Applied Psychology, 88,* 100–112.

Hollingshead, A. B., McGrath, J. E., & O'Connor, K. M. (1993). Group task performance and communication technology: A longitudinal study of computer-mediated versus face-to-face work groups. *Small Group Research, 24,* 307–333.

Holtgraves, T. (2004). Social desirability and self-reports: Testing models of socially desirable responding. *Personality and Social Psychology Bulletin, 30,* 161–172.

Hough, L. M. (1992). The "Big Five" personality variables – construct confusion: Description versus prediction. *Human Performance, 5,* 139–155.

John, O. P., & Robins, R. W. (1993). Determinants of interjudge agreement on personality traits: The Big Five domains, observability, evaluativeness, and the unique perspective of the self. *Journal of Personality, 61,* 521–551.

Judge, T. A., Bono, J. E., Ilies, R., & Gerhardt, M. W. (2002a). Personality and leadership: A qualitative and quantitative review. *Journal of Applied Psychology, 87,* 765–780.

Judge, T. A., Heller, D., & Mount, M. K. (2002b). Five-factor model of personality and job satisfaction: A meta-analysis. *Journal of Applied Psychology, 87,* 530–541.

Judge, T. A., & Ilies, R. (2002). Relationship of personality to performance motivation: A meta-analytic review. *Journal of Applied Psychology, 87,* 797–807.

Kautz, T., Heckman, J. J., Diris, R., ter Weel, B., & Borghans, L. (2014). *Fostering and measuring skills: Improving cognitive and noncognitive skills to promote lifetime success.* Paris, France: Organization for Economic Cooperation and Development.

Kenrick, D., & Funder, D. (1988). Profiting from controversy: Lessons from the person-situation debate. *American Psychologist, 43,* 23–34.

Kerr, D. S., & Murthy, U. S. (2009). Beyond brainstorming: The effectiveness of computer-mediated communication for convergence and negotiation tasks. *International Journal of Accounting Information Systems, 10,* 245–262.

King, G., Murray, C. J., Salomon, J. A., & Tandon, A. (2004). Enhancing the validity and cross-cultural comparability of measurement in survey research. *American Political Science Review, 98,* 191–207.

Kretschmer, E. (1925). *Physique and character.* New York, NY: Harcourt Brace Jovanovich.

Kyllonen, P. C., Lipnevich, A. A., Burrus, J., & Roberts, R. D. (2014). *Personality, motivation, and college readiness: A prospectus for assessment and development* (ETS Research Report No. RR-14-06). Princeton, NJ: Educational Testing Service. 10.1002/ets2.12004

Laughlin, P. R. (1980). Social combination processes of cooperative problem-solving groups on verbal intellective tasks. In M. Fishbein (Ed.), *Progress in social psychology* (Vol. 1, pp. 127–155). Hillsdale, NJ: Erlbaum.

Leising, D., & Borkenau, P. (2011). Person perception, dispositional inferences, and social judgment. In L. M. Horowitz & S. Strack (Eds.), *Handbook of interpersonal psychology: Theory, research, assessment, and therapeutic interventions* (pp. 157–170). Hoboken, NJ: Wiley.

Liberman, V., Samuels, S. M., & Ross, L. (2004). The name of the game: Predictive power of reputations versus situational labels in determining prisoner's dilemma game moves. *Personality and Social Psychology Bulletin, 30,* 1175–1185.

Lonner, W. J. (1981). Issues in testing and assessing in cross-cultural counseling. *The Counseling Psychologist, 13,* 599–614.

Markus, H. (1983). Self-knowledge: An expanded view. *Journal of Personality, 51,* 543–565.

Matsumoto, D. (2007). Culture, context, and behavior. *Journal of Personality, 75,* 1285–1320.

McGrath, J. E. (1984). *Groups: Interaction and performance.* Englewood Cliffs, NJ: Prentice-Hall.

Messick, S. (1979). Potential uses of noncognitive measurement in education. *Journal of Educational Psychology, 71,* 281–292.

Mischel, W. (1968). *Personality and assessment.* New York, NY: Wiley.

Mischel, W., & Shoda, Y. (2008). Toward a unified theory of personality: Integrating dispositions and processing dynamics within the cognitive–affective processing system. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality: Theory and research* (3rd ed., pp. 208–241). New York, NY: Guilford Press.

Motowidlo, S. J., & Burnett, J. R. (1995). Aural and visual sources of validity in structured employment interviews. *Organizational Behavior and Human Decision Processes, 61,* 239–249.

Motowidlo, S. J., & Kell, H. J. (2013). Job performance. In N. W. Schmitt & S. Highhouse (Eds.), *Handbook of industrial and organizational psychology* (2nd ed., pp. 82–103). Hoboken, NJ: Wiley.

Mount, M. K., Barrick, M. R., & Stewart, G. L. (1998). Five-factor model of personality and performance in jobs involving interpersonal interactions. *Human Performance, 11,* 145–165.

Mount, M. K., Barrick, M. R., & Strauss, J. P. (1994). Validity of observer ratings of the big five personality factors. *Journal of Applied Psychology, 79,* 272–280.

Naim, I., Tanveer, I., Gildea, D., & Hoque, M. E. (2015, May). *Automated prediction of job interview performance: The role of what you say and how you say it*. Paper presented at the 11th IEEE International Conference on Automated Face and Gesture Recognition, Ljubljana, Slovenia.

Norman, W. T. (1963). Toward an adequate taxonomy of personality attributes: Replicated factor structure in peer nomination personality ratings. *Journal of Abnormal and Social Psychology, 66,* 574–583.

Oh, I. S., Wang, G., & Mount, M. K. (2011). Validity of observer ratings of the five-factor model of personality traits: A meta-analysis. *Journal of Applied Psychology, 96,* 762–773.

Olguín-Olguín, D. (2007, May). *Sociometric badges: Wearable technology for measuring human behavior* (Unpublished master's thesis). Massachusetts Institute of Technology, Cambridge.

Olguín-Olguín, D., Gloor, P. A., & Pentland, A. S. (2009). Capturing individual and group behavior with wearable sensors. In *Proceedings of the 2009 AAAI Spring Symposium on Human Behavior Modeling* (Vol. 9).

Olguín-Olguín, D., & Pentland, A. (2010a, February). Assessing group performance from collective behavior. Paper presented at the CSCW Workshop on Collective Intelligence in Organizations, Savannah, GA.

Olguín-Olguín, D., & Pentland, A. (2010b). Sensor-based organisational design and engineering. *International Journal of Organisational Design and Engineering, 1,* 69–97.

Organ, D. W., & Ryan, K. (1995). A meta-analytic review of attitudinal and dispositional predictors of organizational citizenship behavior. *Personnel Psychology, 48,* 775–802.

Passini, F. T., & Norman, W. T. (1966). A universal conception of personality structure? *Journal of Personality and Social Psychology, 4,* 44–49.

Paterson, D. G. (1930). *Physique and intellect*. New York, NY: Appleton-Century-Crofts.

Paunonen, S. V., & Ashton, M. C. (1998). The structured assessment of personality across cultures. *Journal of Cross-Cultural Psychology, 29,* 150–170.

Pentland, A. (2006). Automatic mapping and modeling of human networks. *Physica A, 378,* 59–67.

Pentland, A. (2008). *Honest signals: How they shape our world*. Cambridge, MA: MIT Press.

Pervin, L. A. (1994). A critical analysis of current trait theory. *Psychological Inquiry, 5,* 103–113.

Rauthmann, J. F., Seubert, C. T., Sachse, P., & Furtner, M. R. (2012). Eyes as windows to the soul: Gazing behavior is related to personality. *Journal of Research in Personality, 46,* 147–156.

Roberts, B. W., Kuncel, N. R., Shiner, R., Caspi, A., & Goldberg, L. R. (2007). The power of personality: The comparative validity of personality traits, socioeconomic status, and cognitive ability for predicting important life outcomes. *Perspectives on Psychological Science, 2,* 313–345.

Roberts, B. W., & Mroczek, D. (2008). Personality trait change in adulthood. *Current Directions in Psychological Science, 17,* 31–35.

Rodgers, J. L., & Nicewander, W. A. (1988). Thirteen ways to look at the correlation coefficient. *The American Statistician, 42,* 59–66.

Shaffer, J. A., & Postlethwaite, B. E. (2012). A matter of context: A meta-analytic investigation of the relative validity of contextualized and noncontextualized personality measures. *Personnel Psychology, 65,* 445–494.

Shaw, M. (1981). *Group dynamics: The psychology of small groups*. New York, NY: McGraw-Hill.

Solano-Flores, G., & Nelson-Barber, S. (2001). On the cultural validity of science assessments. *Journal of Research in Science Teaching, 38,* 553–573.

Stark, S., Chernyshenko, O. S., Chan, K. Y., Lee, W. C., & Drasgow, F. (2001). Effects of the testing situation on item responding: Cause for concern. *Journal of Applied Psychology, 86,* 943–953.

Steiner, I. D. (1972). *Group process and productivity*. New York, NY: Academic Press.

Straus, S. G. (1999). Testing a typology of tasks: An empirical validation of McGrath's (1984) group task circumplex. *Small Group Research, 30,* 166–187.

Straus, S. G., & McGrath, J. E. (1994). Does the medium matter? The interaction of task type and technology on group performance and member reactions. *Journal of Applied Psychology, 79,* 87–97.

Sudman, S., Bradburn, N., & Schwarz, N. (1996). *Thinking about answers: The Application of cognitive processes to survey methodology*. San Francisco, CA: Jossey-Bass.

Thomas, A., & Chess, S. (1977). *Temperament and development*. New York, NY: Brunner/Mazel.

Waber, B. N., Olguín-Olguín, D., Kim, T., & Pentland, A. (2008, August). *Understanding organizational behavior with wearable sensing technology*. Paper presented at the Academy of Management Annual Conference, Anaheim, CA.

Watson, R. T., DeSanctis, G., & Poole, M. S. (1988). Using a GDSS to facilitate group consensus: Some intended and unintended consequences. *MIS Quarterly, 12,* 463–478.

Wiley, D. (1991). Test validity and invalidity reconsidered. In R. Snow & D. Wiley (Eds.), *Improving inquiry in the social sciences* (pp. 75–107). Hillsdale, NJ: Erlbaum.

Wilt, J., & Revelle, W. (2015). Affect, behaviour, cognition and desire in the Big Five: An analysis of item content and structure. *European Journal of Personality, 29,* 478–497.

Woolley, A. W., Chabris, C. F., Pentland, A., Hashmi, N., & Malone, T. S. (2010). Evidence for a collective intelligence factor in the performance of human groups. *Science, 330,* 686–688.

Wu, L., Waber, B. N., Aral, S., Brynjolfsson, E., & Pentland, A. (2008, December). Mining face-to-face interaction networks using sociometric badges: Predicting productivity in an IT configuration task. In *Proceedings of the International Conference on Information Systems* (pp. 1–19). Paris, France.

Zillig, L. M. P., Hemenover, S. H., & Dienstbier, R. A. (2002). What do we assess when we assess a Big 5 trait? A content analysis of the affective, behavioral, and cognitive processes represented in Big 5 personality inventories. *Personality and Social Psychology Bulletin, 28,* 847–858.

Suggested citation:

Klafehn, J., Kell, H., Andrews, J., Barnwell, P., & Khan, S. (2016). *We are what we repeatedly do—But in what context? The role of situational factors in assessing personality via nonverbal behavior.* (Research Report No. RR-16-28). Princeton, NJ: Educational Testing Service. http://dx.doi.org/10.1002/ets2.12114

Action Editor: Jim Carlson

Reviewers: Michelle Martin-Raugh and Blair Lehman

Find other ETS-published reports by searching the ETS ReSEARCHER database at http://search.ets.org/researcher/