IPAC
INTERNATIONAL PERSONNEL ASSESSMENT COUNCIL

ASSESSMENT COUNCIL
NEWS

President's Message

Matisha Montgomery, IPAC President

As I write this message, it has already been six weeks since we closed out the 2018 IPAC Conference in Old Town Alexandria. That is unbelievable to me since I am still "a-twitter" from the energy you all shared with me over those three days. I have so much personal and professional pride in what we produced for you this year and in the close-knit family we created in the process. My gratitude to the many IPAC member-volunteers that helped bring this conference together isn't sufficient. I

*IPAC Past President **Mary Ann Haskins** (left) with IPAC President **Matisha Montgomery** (right).*

hope those of you who were able to attend will take a moment to share your appreciation and favorite moments as well. Help me give kudos to everyone involved in the planning and share what you loved about this year's conference #IPAC2018 by tweeting us @ipacweb and @matishamo.

Whether you were able to attend this year or not, you can read a recap of the 2018 conference on page 3. Our conference chair Kathy Stewart took on a big job leading the conference planning committee and approached each task with grace, humility, a sense of humor, and a smile. Kathy, thank you for being the most awesome you! Our program chair, Julie Weintraub, brought us a unique lineup of keynote speakers, a diverse group of pre-conference workshops, and a long list of concurrent sessions that catered to early career and tenured professionals alike. I'm so excited she is continuing in her program chair role in 2019. Choni Gurira, host chair, and Emily Steinau,

October 2018

social chair, made sure all of us felt like their invited guests at the Westin and at our social events. Lindsay Northon served as this year's sponsor chair helping to ensure IPAC brings you a group of vendor partners that can help meet your organization's goals but also help keep your costs low to attend the conference. There are many others that deserve my heartfelt thank you as well: Alexis Avery, Bharati Belwalkar, Nathan Carter, Stephanie Cramer, Mary Ann Haskins, Martha Hennen, Dennis Joiner, Christopher Nye, Erin Smith, and James Wilcox. Thank you to this amazing group of people! Y'all knocked my socks off!

The IPAC annual conference continues to live up to our core values: Community, Learning, and Practice. The Board is working in support of these values by carrying out our 2018-2020 strategic plan. If you would like to help support one of our strategic teams, please contact me or the team lead. Some highlights currently underway:

• **Establish regional chapters** (e.g., GLEAN). A new chapter in the south plains / south central states (Texas, Oklahoma, Arkansas, Louisiana area) is currently forming. Contact Liz Reed at ereed1@columbus.gov

• **Develop additional content and learning opportunities**. Would you or your organization like to deliver or support a webinar, present a training, or publish an Monograph? Contact Dennis Doverspike at dennisdoverspike@gmail.com

• **Enhance and redesign IPAC website**. Are you technically savvy and have experience with website management? Contact Ben Porr at elcomnet@ipacweb.org

Now we will also start to transition our time and attention to the fun that we will be bringing to Minneapolis, MN in 2019. Be sure to save the date and make plans to join us at the Embassy Suites Downtown Minneapolis, July 14-17, 2019. If you are interested in assisting with the conference planning committee please email me at President@ipacweb.org or Marty Alber (President-Elect) at Marty.Alber@pbjcal.org.

I look forward to reuniting or meeting you in Minneapolis! #IPAC2019

IPAC 2018 - Conference Highlights

Save the Date IPAC 2019!
July 14-17, 2019 • Embassy Suites Downtown
Minneapolis, MN • #IPAC2019

Conference Chair **Kathy Stewart** *(left) and Student Paper Award Winner* **Jacob Bradburn** *(right)*

A group congratulating Bemis Awardee **Marianne Tonjes** *(second from left) at the Monday Night Social*

Dennis Doverspike, *left, and* **Eleni Lobeni**, *right, Innovations in Assessment Award representative*

From left, **Emilee Tison**, **Kristen Pryor**, **Jeff Cucina**, **Kayo Sady**

IPAC's 43rd annual conference held July 29 – August 1, 2018 in Old Town Alexandria, VA generated a great deal of energy, enthusiasm, and new connections. We had 231 attendees join us from across the country, and even some who came from outside the United States.

The charm of Old Town provided attendees with plenty of opportunities to explore just steps from the conference hotel, through the city's cobblestone streets, historic attractions, and superb restaurants and shopping.

This year's program included a diverse group of presenters and topic areas. Attendees chose from a myriad of learning opportunities, featuring 60 concurrent sessions, workshops, keynotes, and posters. The program provided information for attendees with all levels of knowledge of the various topic areas (basic to advanced) and all pre-conference workshops were approved for professional development credits through SHRM and HRCI. The presentations are always free for members and will be available soon at http://www.ipacweb.org/Conference-2018-Presentations

Social activities are one of the cornerstones of the IPAC conference, showing the true meaning of "IPAC family." This year, attendance at all social events exceeded expectations, as did the conversation, food, drinks, and laughs.

The President's Reception celebrated the beginning of the conference and gave attendees an opportunity to connect and reconnect with each other and meet with the exhibiting sponsors.

The Monday Night Social Event at Virtue Feed and Grain proved to be a memorable evening, with attendees toasting to the conference amidst the backdrop of a beautiful historic building right in the heart of Old Town.

Some attendees joined us in traveling to DC to partner in a joint-happy hour with our friends from the Personnel Testing Council of Metropolitan Washington (PTCMW), while others stuck around Old Town and took advantage of the wonderful restaurants through our annual group dinners.

Finally, IPAC's Hospitality Suite was open evenings and late nights and provided a laid-back atmosphere for attendees to relax, connect, and have some fun with friends old and new.

*Keynote speaker **John C. Scott***

Some first time attendees getting an IPAC Welcome at the President's Reception

Miss the conference? See more on our Facebook Page.

Members of the IPAC community were presented with well-deserved awards, including the Bemis Award to Marianne Tonjes, the Student Paper Award to Jacob Bradburn (University of Michigan), the Clyde J. Lindley Exemplary Service Award to Mary Ann Haskins (City of Johns Creek), and the Innovations in Assessment Award to Eleni Lobene, Alexander Stemer, Tedd Shapiro, Tara Johnson, Sarah Meeks, Michelle Heikkila, Jeff Ryer (all from Aon Hewitt) and Jon Exline and Jonathan Neff (Anheuser-Busch). Be sure to look for Jacob Bradburn's award winning paper in the December edition of the ACN.

We hope you'll join us next year July 14-17, 2019 in Minneapolis!

IPAC would like to thank all the Sponsors & Exhibitors

VIRTUAL JOB TRYOUT®

Award-winning by design

Virtual Job Tryout® is a job simulation recruiting technology that engages, educates, & evaluates while delivering an award-winning candidate experience.

Virtual Job Tryout® is now offered in **three levels** of implementation:

NEW!

- **Standard**—Ready-to-use technology
- **Tailored**—Branded to your organization
- **Custom**—Realistic preview specific to your job

Advances in Assessment Technology: Continuous Rating Assessment

Kristen Hassenkamp, University of Akron

Change happens in the blink of an eye. Imprecise measurement of change can undermine research and hide significant results. As the field of psychology advances into an age of fast-paced technology, psychological assessment methods are incorporating the benefits offered by technology to solve various assessment problems. One issue with traditional assessments is that they are often static, which is not a problem when measuring a highly stable trait. However, when measuring a dynamic characteristic of an individual, there is a need for rapid, on-demand assessments. Continuous rating assessments (CRA) are perfectly situated to measure dynamic, fast-changing constructs over a short period of time. For example, the act of responding to difficult customer service interactions in a dispatch center is one where we would want to take repeated measures across calls and even within calls, in order to determine how the performance and the emotional reactions of the employee change over time. Using CRA can help practitioners better understand the underlying movements of the change they observe by revealing patterns that may be unexpected.

When examining changes over time, past issues have revolved around when to take measurements. If too much time passes after a change-inducing event, the effect may have faded, with data showing no difference. Some constructs may have a delay in change, where measuring too soon after an event will also show no difference in a variable.

Continuous rating assessments offer a way for researchers to collect data that goes beyond answering simple questions such as what is someone's average level of satisfaction or has a change in stress level occurred over a two week period. Because of the continuous nature of CRA measurement, there are no significant gaps of time where the variable is not being measured. This provides information on when, why, and how changes occur.

Traditional methods ask participants to deliver a global rating post-event, masking within-person changes due to averaging across the event. In comparison to diary methods, or other assessments where individuals are asked to report variables repeatedly across some length of time, CRA can also show the rate of change, why the construct changes, and allow comparisons in how individuals change.

How Does It Work?
In an event such as a difficult customer service interaction or during an interview, people will react to and change their psychological state throughout the event. In the call center example, CRA would ask participants to rate their affect levels continuously during the interaction with a difficult customer. Instead of revealing simply that difficult customers cause employees to experience more negative affect, CRA would be able to show how affect changed (gradually or in large drops) and in response to what part of the conversation.

To collect CRA data at the individual level, a dial, a sliding knob, or a computerized scale is used that allows individuals to adjust their ratings at any time; this can be easily programmed using a variety of computer software. The computer then records information multiple times per second, detecting any meaningful change. Ratings can be gathered during the event, but most researchers find this can disrupt the experiencing of the event (i.e., responding to an angry customer), so another option is to record the event and replay it at a later time, i.e., an after-action-review (AAR). This cued-recall CRA asks participants to rate how they felt during the event as the employee relives it.

Validity
Empirical research of the validity of continuous rating assessments has been promising. Several affect studies

examined covariance between the CRA and other ratings. A study in 1985 by Gottman and Levenson found high levels of agreement between observer ratings and self-ratings of affect. Physiological responses such as heart rate showed similar patterns while re-watching the event and during the event itself, supporting the cued-recall validity. A study in 2005 by Mauss and colleagues compared CRA ratings of sadness and amusement to trained observers' ratings, as well as physiological responses. The average correlation between continuous ratings and post-event ratings was .51 for amusement and .82 for sadness, with correlations of .67 for amusement and .62 for sadness in a cued-recall CRA design.

Some Considerations

If you are interested in using CRA, it is important to consider whether construct or characteristic of interest is amenable to being measured by this method. Continuous rating can be exhausting for participants, so the events under examination should be relatively brief, usually under 10 minutes. As a practical consideration, if you want to get ratings for many different variables, the length of the event being rated should be adjusted to reflect the number of measures taken or responses necessary the participant. For example, if 3 variables are to be measured, the event should of no more than three minutes. If measuring only 1 variable, the event can be longer. Due to these time constraints, the variables of interest should be dynamic enough that changes are likely to occur within the ten-minute time frame.

Benefits and Applications

Some benefits to this method are its ability to show within-event, within-person changes, giving a more nuanced understanding of change. Identifying different patterns of change between participants can shed light on stable traits that influence how change occurs. For affect research, emotional intelligence may influence differences between individuals. As many studies do not address dynamic change in the variables being studied, CRA supports theories as well as builds theories, now that it is possible to measure within-event rate of change. Much research using CRA has been done in the affect literature, but I hope this brief summary encourages practitioners to consider using the technique in studies of recruitment, performance appraisal, and leadership.

For additional information, please consider:

Gabriel, A. S., & Diefendorff, J. M. (2015). Emotional labor dynamics: A momentary approach. *Academy of Management Journal,* 58, 1804-1825.

Gabriel, A. S., Diefendorff, J. M., Bennett, A. A., & Sloan, M. D. (2017). It's about time: The promise of continuous rating assessments for the organizational sciences. *Organizational Research Methods*, 20, 32-60.

Gottman, J. M., Levenson, R. W. (1985). A valid procedure for obtaining self-report of affect in marital interaction. *Journal of Consulting and Clinical Psychology*, 53, 151–160.

Mauss, I. B., Levenson, R. W., McCarter, L., Wilhelm, F. H., Gross, J. J. (2005). The tie that binds? Coherence among emotion experience, behavior, and physiology. *Emotion*, 5, 175–190.

"*IPMA-HR provides excellent tests at a very reasonable price! They have a dependable and dedicated team that is always willing to answer questions in a timely manner and provide exceptional customer service. They are great to work with! The City of Owosso has been a satisfied customer for over three years.*"
— Jessica Unangst, Director of Human Resources, City of Owosso

POLICE FIRE EMERGENCY COMMUNICATIONS CENTER CORRECTIONS ADMIN SUPPORT

Gain more information and insight than interviews and applications alone can provide. With IPMA-HR's assessment products, you can hire and promote with confidence knowing you're using the most reliable, comprehensive, and affordable public safety assessment products in the industry.

ASSESSMENT PRODUCTS
HIRE & PROMOTE WITH CONFIDENCE
DEVELOPED BY IPMA-HR

learn MORE!

Visit us online
ipma-hr.org/Assessment

Or call
1-800-381-TEST (8378)

2018 Clyde J. Lindley Exemplary Service Award Recipient: Mary Ann Haskins

*2018 IPAC President **Matisha Montgomery** (left) with Clyde J Lindley award recipient **Mary Ann Haskins** (right).*

The original IPAC Exemplary Service award was established by the IPMA Assessment Council (IPMAAC) in 1991, in conjunction with IPMAAC's 15th anniversary as a way to recognize members for continuously and generously contributing their time, energy, and talents to IPMAAC. The Award was renamed the Clyde J. Lindley Exemplary Service Award in 1999 in honor of one of IPMAAC's founding members, the late Clyde J. Lindley.

There are few as deserving as this year's recipient, Mary Ann Haskins. Mary Ann has gone well above and beyond the call of duty for years. She has volunteered countless hours of her personal time, staff from her team in Johns Creek, GA, her prior experiences and expertise, her energy and passion for this community, and of course, we can't fail to mention, the margarita machine.

As President-Elect, President, and now Past President, there is an expectation that she is giving to the organization. But what Mary Ann gives is something much more. Mary Ann has partnered with so many across the IPAC community to ensure we establish and build upon a strong foundation for our current and future business practices. She has consistently served in multiple roles while also leading the Board. She has strengthened IPAC's relationship with partner organizations such as IPMA-HR and PTC/MW to keep IPAC always moving forward.

Thank you, Mary Ann Haskins, for your continued service and dedication to IPAC and its ideals. Congratulations!

14TH ANNUAL SIOP LEADING EDGE CONSORTIUM

HIGHPOTENTIAL:
IDENTIFYING, DEVELOPING & RETAINING FUTURE LEADERS

OCTOBER 19–20, 2018
RENAISSANCE BALTIMORE HARBORPLACE HOTEL
OPTIONAL WORKSHOPS ON OCTOBER 18

Top thought leaders together with seasoned practitioners who are knowledgeable and experienced in all aspects of high potentials.

FOR MORE INFORMATION OR TO REGISTER, VISIT MY.SIOP.ORG/MEETINGS/LEADING-EDGE-CONSORTIUM

Testing Programs and the Epic Quest for Suitable Alternatives

Kayo Sady, Ph.D., Associate Principal Consultant, DCI Consulting Group, Inc.,
and Emilee Tison, Ph.D., Associate Principal Consultant, DCI Consulting Group, Inc.

Abstract: This article introduces a series of articles to focus on the search for suitable alternative selection procedures, as prescribed by the Uniform Guidelines on Employee Selection Procedures (1978). We present background information on the principle of suitable alternatives and review aspects of the Supreme Court ruling in Albermarle Paper v. Moody (1975) that related to the search and adoption of suitable alternatives.

Keywords: Suitable Alternative, Albermarle, Job-relatedness

If a testing program produces differential pass rates between protected class subgroups (e.g., sex, race/ethnicity), that give rise to a legitimate adverse impact claim, employers can share validation research to defend against the claim. A technical report that demonstrates the job-relatedness of the program typically provides the basis for defending the program. In response, plaintiffs may try to rebut the validity evidence by arguing that an equally valid 'suitable alternative' with smaller subgroup differences should have been implemented instead. The question of whether a proffered substitute is a legitimate 'suitable alternative,' however, has stymied progress on no small number of employment discrimination cases. In this and follow-up articles, we highlight features that should be evaluated in determining whether a proffered replacement is truly a 'suitable alternative.' The remainder of this installment provides background information for follow-up articles, which will focus on the following questions:

• What characteristics of an alternative assessment qualify it as 'suitable'?
• What level of burden is required of the employer to seek and identify an alternative?
• Assessment expertise aside, what do the courts say?

The concept of suitable alternatives emerged from the Equal Employment Opportunity Commission (EEOC) Guidelines (EEOC Guidelines), which were relied on in two landmark Supreme Court cases: Griggs v Duke Power Co. (1971) and Albermarle Paper Co. v Moody (1975). In Albermarle Paper Co. v Moody (1975), several employment practices were contested, only one of which (the testing program for selecting skilled tradespeople) is relevant to the topic of suitable alternatives. The testing program required applicants to have a high school diploma, pass a non-verbal intelligence test, and pass the Wonderlic Personnel Test in order to be hired into skilled progression roles. Plaintiffs (African-American applicants) challenged the job-relatedness of the testing program and argued that it was discriminatory based on the pass rate differences between African-American candidates and White applicants.

The District Court was favorable to the employer, ruling that:
The personnel tests administered at the plant have undergone validation studies and have been proven to be job related. The defendants have carried the burden of proof in proving that these tests are `necessary for the safe and efficient operation of the business' and are, therefore, permitted by the Act. However, the high school education requirement used in conjunction with the testing requirements is unlawful in that the personnel tests alone are adequate to measure the mental ability and reading skills required for the job classifications.
Although the ruling endorsed Abermarle's use of the pre-employment tests, the reasoning behind the court's striking of the high school diploma requirement is inherently consistent with a proposed 'suitable alternative':

• The two tests provided an indication of whether applicants had sufficient cognitive ability and reading skill to perform the required tasks of the jobs;
• The addition of the high school diploma criteria added no value and increased pass rate differences between African-American candidates and White candidates.

The appellate court overturned the district court's determination that the two employment tests had been sufficiently demonstrated to be job-related, noting that the EEOC Guidelines were "entitled to great deference" in evaluating the job-relatedness of employment tests; the Supreme Court affirmed the appellate court ruling and reasoning in the majority opinion, delivered by Justice Stewart. The court's ruling specifically addressed the opportunity for plaintiffs to present alternatives:

…the respondents have not until today been specifically apprised of their opportunity to present evidence that even validated tests might be a "pretext" for discrimination in light of alternative selection procedures available to the Company.

In 1978, the Uniform Guidelines on Employee Selection Procedures (UGESP) codified guidance on the topic of suitable alternatives:

Where two or more selection procedures are available which serve the users legitimate interest in efficient and trustworthy workmanship, and which are substantially equally valid for a given purpose, the user should use the procedure which has been demonstrated to have the lesser adverse impact. Whenever the user is shown an alternative selection procedure with evidence of less adverse impact and substantial evidence of validity for the same job in similar circumstances, the user should investigate it to determine the appropriateness of using or validating it in accord with these guidelines. [emphasis added]

This guidance indicates that if two selection procedures are equal in their job-relatedness and accomplishment of the assessment's objectives, but have different levels of adverse impact, the procedure associated with lower adverse impact should be used. Such guidance is part and parcel with that provided by the EEOC Guidelines referenced earlier. Following the guidance can lead to the use of selection procedures that maximize the validity of selection decisions while minimizing adverse impact against protected classes. However, this guidance stops short of addressing what constitutes an alternative (i.e. what employers should be comparing in terms of validity and adverse impact) and what may be suitable, two very important details to consider when comparing testing programs.

The requirement to seek and evaluate suitable alternative selection procedures can be a complex issue in practice. Unfortunately, guidance on the topic is relatively scarce, and there is widespread misinterpretation of what constitutes a suitable alternative and what efforts are required to seek out a suitable alternative. Interpretation of the Uniform Guidelines and research coming out of the I-O psychology world highlights existing points of controversy or disagreement among experts. Case law is generally just as mixed in opinion. In the follow-up installments of this series, we hope to present cogent summaries of some details that bear upon the evaluation of suitable alternatives. Until next time, may you find both work and life to be reliable and valid.

References & Cases

Albemarle Paper Co. v. Moody, 422 U.S. 405 (1975)
Griggs v. Duke Power Co., 401 U.S. 424 (1971)
Equal Employment Opportunity Commission, Civil Service Commission, Department of Labor, & Department of Justice. (1978). Uniform guidelines on employee selectin procedures. *Federal Register*, 43.

Cross-Cultural Generalization of a Multidimensional Pairwise Preference Inventory: Methodological Considerations and Empirical Findings

Christopher R. Huber, Human Resources Research Organization;
John Capman, Anthony Boyce, and Eleni Lobene, Aon

Abstract: Organizations have become increasingly interested in using personality tests on a global scale. We discuss a relatively new type of personality test designed to resist faking, which presents both benefits and challenges as a cross-cultural personality measure. First, we demonstrate the need to obtain culture-specific estimates of social desirability and discuss a simpler approach to do this. Second, we discuss a simpler method to examine the test's functioning in new cultures. Using this approach can facilitate the development of a globally relevant test that puts different cultures on a level playing field.

Keywords: forced choice, cross-cultural, personality

Spurred by two meta-analyses (Barrick & Mount, 1991; Tett, Jackson, & Rothstein, 1991), personnel psychology has seen a growing interest in personality over the past few decades. Personality traits have often been measured using a classical test theory approach where scores on single-statement items are added together to produce a total score (e.g., Likert scales). However, researchers have argued that this approach produces suboptimal measures, citing limitations such as test length, ease of faking, and lack of flexibility (e.g., Stark, Chernyshenko, Drasgow & White, 2012). Recent research has applied item response theory (IRT) to binary forced choice measures in order to alleviate these problems while still allowing for normative comparisons between individuals (Stark, Chernyshenko, & Drasgow, 2005). The resulting multidimensional pairwise preference (MDPP) measures have exhibited desirable properties, including decreased fakability, increased versatility, and potentially increased validity (Boyce, Mead, & Conway, 2014; Salgado & Táuriz, 2014; Stark et al., 2005; Stark et al., 2012).

Given the benefits of these measures, there has been increasing interest in expanding their use for personnel selection and development. However, the development of MDPP inventories presents novel challenges for applied psychologists. One significant challenge is adapting these inventories for use in different cultures, which can be complex for a variety of reasons. One major concern is that the social desirability of personality traits and items may vary across cultures. This possibility is problematic because MDPP testing relies on accurate social desirability estimates to create faking-resistant items. Furthermore, establishing unique social desirability parameters for each culture can be difficult and costly. A second issue is that the structure of MDPP items restricts the use of traditional measurement invariance analyses. In particular, the adaptive nature of many MDPP assessments precludes administration of the same items to all test takers. As a result, it is logistically difficult to analyze measurement equivalence across different cultural groups.

The purpose of this paper is to explore these issues, provide solutions, and present related empirical findings. To do so, we discuss the cross-cultural generalization of the Adaptive Employee Personality Test (ADEPT-15; Boyce, Conway, & Caputo, 2016), the first commercially available MDPP inventory for personnel selection and winner of the 2016 SIOP M. Scott Myers Award. First, we provide evidence that social desirability parameters can be estimated using relatively few expert raters. Second, we apply this methodology across several cultures and provide evidence for the importance of obtaining culture-specific estimates. Finally, we describe an alternative procedure for analyzing measurement invariance with adaptive MDPP tests.

MDPP Testing

The resurgence of forced choice measures arose in part from Stark's (2002) application of IRT to extract normative data from multidimensional statement pairs. These forced choice items require examinees to choose which of two statements describes them better. The paired statements typically reflect different dimensions of personality and are matched on social desirability (see Figure 1 for examples). Using traditional scoring methods, such items would only facilitate within-person comparisons (e.g., "John is more extraverted than he is agreeable"). However, Monte Carlo simulations have shown that MDPP tests scored using IRT can produce normative scores by including some unidimensional statement pairs (Stark et al., 2005). In addition, normative scores can be extracted using as few as one or two unidimensional items per dimension.

Social Desirability Estimation

The faking-resistance of MDPP items hinges on obscuring the relative desirability of different response options. To accomplish this, items are constructed by pairing personality statements with similar levels of social desirability. Thus, it is essential to produce accurate social desirability estimates for each statement. There are currently two main methods for estimating social desirability parameters (Drasgow et al., 2012; Stark et al., 2005). The first approach is to estimate social desirability using a directed faking study. In this approach, participants complete individual personality statements with instructions to portray themselves as good employees. Social desirability parameters are then computed using mean endorsement rates from this sample.

However, this method has at least two significant limitations. First, Stark et al. (2005) found that directed faking social desirability estimates were correlated with the latent trait level of each statement. For example, a statement reflecting a high level of conscientiousness would also be likely to have a high social desirability rating. This presents a logistical challenge because paired statements should have different locations on the latent trait continuum in order to maximize item information. An additional consideration for cross-cultural research is that conducting directed faking studies in several cultures can be logistically infeasible. For adaptive MDPP tests like ADEPT-15, the total statement pool can consist of well over 1,000 statements (Drasgow et al., 2012; Boyce, Conway, Caputo, & Huber, 2015). Since each statement receives responses from a substantial number of directed fakers, replicating such an endeavor several times can be costly.

A second approach to social desirability estimation is to simply ask judges to rate the social desirability of each statement. Stark et al. (2005) recommended this method for future research in order to reduce correlations between latent trait levels and social desirability ratings. However, it is unclear what effect this would actually have in practice. It is certainly likely that directed fakers endorse statements based on their trait levels, but judges may assign social desirability ratings in a similar manner. As such, social desirability and trait level estimates may be correlated regardless of the estimation method employed.

Unfortunately, research comparing the two methods is limited. Drasgow et al. (2012) conducted one such comparison using two sets of Army recruits as raters. With 30-40 ratings per statement, they found a correlation of .87 between the two methods. This suggests that fakers and judges do provide similar ratings of social desirability. As a result, using judges is unlikely to substantially reduce the correlation between social desirability and trait level estimates. Furthermore, collecting ratings from judges across several cultures presents the same logistical challenges as conducting a directed faking study.

In order to reduce this burden, we investigated the possibility of using a relatively small sample of judges to establish culture-specific social desirability ratings. To do so, we compared ratings obtained from a large directed faking study to those provided by two I/O psychologists. For the directed faking study, we recruited a sample of 1,841 American workers from Amazon's Mechanical Turk (MTurk). Each participant responded to a subset of the total statement pool, resulting in a sample size of approximately 70 responses per statement. Like Drasgow et al. (2012), we found a high correlation between the two types of social desirability rating, $r = .82$, $p < .01$. Thus, it appears that useful social desirability ratings can be obtained from a small number of expert judges. This drastically increases the feasibility of collecting multiple sets of culture-specific ratings.

Measurement Invariance Analysis

Another challenge associated with cross-cultural transportability is demonstrating that a test functions the same way in different cultural groups. Typically this can be done using multi-group confirmatory factor analysis (MGCFA), which tests the equivalence of a test's factor structure across groups. This is accomplished by comparing a se-

ries of CFA models with increasingly rigid equality constraints (Cheung & Resvold, 2002). The simplest model, configural invariance, requires that the same items load onto the same constructs across groups. The next model, construct-level metric invariance, constrains the overall strength of the factor loadings to be equal across groups. Next, scalar invariance adds an equality constraint to the item intercepts. In other words, the "values of each item corresponding to the zero value of the underlying construct" must be equal (Cheung & Rensvold, 2002, p. 237). Finally, mean invariance constrains the latent construct means to be equal. It is particularly useful to establish metric invariance, since this provides evidence that relationships between test scores and external criteria will be invariant across the groups (Bollen, 1989; Cheung & Rensvold, 2002).

Unfortunately, measurement invariance analyses such as MGCFA require statement-level data. This is a significant limitation for MDPP testing, since statements are presented in pairs. The difficulty is increased for adaptive tests, since a given participant can see any number of statement pairs generated by the adaptive algorithm. For example, the ADEPT-15 algorithm can produce over 350,000 unique statement pairs. Given the burden of administering hundreds of thousands of items (or thousands of individual statements) to a large calibration sample, it may be infeasible to conduct traditional invariance analyses across several cultures.

However, it is still useful to demonstrate some degree of cross-cultural measurement equivalence. To do this, we analyzed equivalence at the construct level rather than the item level by treating ADEPT-15 dimension scores as items that loaded onto higher-order factors. Conceptually, this should determine whether the relationships between MDPP dimensions remain the same across groups. This is particularly important for multidimensional forced choice scales since the dimensions are measured simultaneously. Furthermore, this approach allows for the use of existing scale-level data collected via operational use of the assessment.

Study 1 – Social Desirability

Participants
Our rater dataset included 140 graduate students in psychology, psychometrics, and related disciplines from around the world. Each rater was selected to represent one of 31 cultural groups based on nationality and language (see Table 1), with each group containing a minimum of three raters. Raters were required to be native speakers of the targeted language and to have lived most of their lives in the target country. Raters were predominantly female (67%) with a mean age of 26.9 years (SD = 5.0).

Measures
Participants provided social desirability ratings for the ADEPT-15 statement pool, which consists of 1,470 unidimensional statements that measure 15 dimensions of personality. The first 10 dimensions are based on DeYoung, Quilty, and Peterson's (2007) Big Five Aspect Scales, which measure two lower-level aspects for each Big Five personality trait (see Table 2 for Big Five mappings). The remaining five dimensions are focused on traits not traditionally covered by the Five Factor Model (FFM): (1) Ambition, (2) Awareness, (3) Humility, (4) Mastery, and (5) Power. The Ambition dimension was based on research on the need for achievement (Steers & Braunstein, 1976). The Awareness dimension was developed based on research examining social effectiveness and emotional intelligence (e.g., Joseph & Newman, 2010). The Humility dimension was based on the HEXACO model developed by Lee and Ashton (e.g., Lee & Ashton, 2010). The Mastery dimension was based on VandeWalle's learning goal orientation construct (e.g., VandeWalle, 1997). Finally, Power was based on research on the need for power (Steers & Braunstein, 1976).

Procedure
Participants provided their ratings through a web-based survey. Given the amount of time necessary to rate the full statement pool, they were allowed to complete the task over several days and instructed to take breaks at least every 45 minutes. In order to obtain culture-specific ratings, participants were asked to rate the statements based from the perspective of someone from their own culture. Participants received a $150 Amazon gift certificate as compensation for their time.

Results
We analyzed cross-cultural variance in social desirability at both the construct level and the statement level. Construct-level differences suggest that cultures vary in their valuation of a global personality trait, while statement-level differences suggest that specific statements are perceived differently across cultures. To analyze construct-level variance, we computed average statement scores for each ADEPT-15 dimension and fit a one-way analysis of variance

(ANOVA) with culture as the independent variable. As shown in Table 3, we found significant differences for 12 of the 15 ADEPT-15 dimensions (with two more reaching marginal significance), suggesting that there were significant cross-cultural differences in social desirability. We also found significant differences in overall social desirability ratings obtained by averaging all ADEPT-15 items. This may indicate an overarching response set that produces higher ratings in some countries, regardless of the specific personality dimension.

As a result, it is also possible that between-culture differences in rating patterns could explain the observed differences on individual personality dimensions. If this were the case, cross-cultural variance on a given trait would be purely artifactual rather than substantive. In order to investigate this possibility, we conducted an analysis of covariance (ANCOVA) for each dimension and controlled for overall social desirability ratings. The overall rating covariate reached significance at the .01 level in all cases, suggesting that a general rating tendency did impact between-group differences on individual traits. However, as shown in Table 3, the effect of culture was also significant at the .01 level for all dimensions (including the three for which there was previously no effect). Given this pattern of results, it appears that variance in social desirability ratings is a function of both culture-based rating biases and true differences in social desirability. For the three dimensions with non-significant ANOVAs, the effect of rating biases may have also obscured true variance in social desirability.

Variance in statement level social desirability estimates is even more important for MDPP testing since item formation occurs at this level. Ideally we could use multivariate ANOVA to perform a single variance test for each ADEPT-15 dimension. However, this analysis was not possible given the large number of items per dimension. Instead, we ran individual one-way ANOVAs for each item and summarized the percentage of significant ANOVAs by dimension (see Table 3). Given the large number of significance tests, we expected some tests to reach statistical significance purely by chance. Since we set the type I error rate to .05, approximately five percent of the item-level ANOVAs should reach significance if the null hypothesis is true. However, the actual percentage of significant results was 39.5, with values ranging from 27.1 to 60.2 for individual dimensions. This suggests that certain items do vary in social desirability across cultures, while others do not.

Study 2 – Measurement Invariance

Participants
In order to test the invariance of ADEPT-15 factors across cultures, we analyzed existing data collected for selection purposes. The complete sample consisted of 24,351 applicants for professional, managerial, and executive positions at 22 companies in 11 countries (see Table 4).

Measures
Applicants completed ADEPT-15, a computer adaptive MDPP personality inventory. ADEPT-15 was constructed using IRT parameters from the generalized graded unfolding model (Roberts, Donoghue, & Laughlin, 2000), which is an ideal-point IRT model (see Drasgow, Chernyshenko, & Stark, 2010, for a non-technical overview). The scoring for the assessment is predicated on the multi-unidimensional pairwise preference model developed by Stark et al. (2005). Previous CFA research on the ADEPT-15 dimensions has supported a higher-order factor structure consisting of six work styles (Boyce et al., 2016; see Table 2 for details). For the measurement invariance analysis, we attempted to replicate this factor structure across countries.

Results
We conducted a MGCFA to analyze the invariance of the ADEPT factor structure across 11 countries. In keeping with recommendations from simulation research, we reported three fit indices in addition to chi-square: the Comparative Fit Index (CFI), McDonald's (1989) Non-Centrality Index (NCI), and Gamma hat (Steiger, 1989). These indices produce model comparisons that are independent of model complexity, sample size, and overall model fit, and they also provide non-redundant information (Cheung & Rensvold, 2002; Meade, Johnson, & Braddy, 2006). In order to determine the significance of these metrics, we used the empirical thresholds established by Cheung and Rensvold (2002) for changes in fit between different invariance models.

The MGCFA results are shown in Table 5. The model for metric invariance was the best fitting model, as it constrained the factor loadings across both groups to be equal but did not cause a significant increase in CFI, NCI, or Gamma hat. The difference in model chi-squares was significant, but this was expected in our sample since the chi-square significance test is heavily dependent on sample size. Although full measurement invariance was not supported, the existence of metric invariance confirms that the 6-factor model holds across countries (even though some

differences in mean levels may exist).

Discussion

The purpose of the present study was to investigate a process for generalizing MDPP inventories across cultures and raise awareness about important considerations in cross-cultural research. The results from Study 1 demonstrate the importance of considering cross-cultural differences in social desirability on a few levels. First, they demonstrate that cultures can differ in their valuation of many (if not most) personality traits. Only three ADEPT-15 dimensions did not show significant between-group variance in social desirability, and only one of those failed to reach marginal significance.

Our results also suggest that some cultures produce higher or lower social desirability ratings in general. This may be attributed to cross-cultural differences in one or more response sets (e.g., acquiescence bias), but further research is needed to fully explain this phenomenon. Importantly, the rating artifact did not account for all observed differences in individual dimension ratings. In fact, our results suggest that it sometimes obscured true differences in trait evaluations. Thus, it is important for cross-cultural researchers to consider the effects of response sets on observed results.

At the item level, our results also support the importance of using culture-specific social desirability ratings for MDPP testing. The social desirability of many items varied across cultures, suggesting that simply using ratings from a single culture will produce suboptimal estimates. On the other hand, it appears that useful estimates can be obtained from as few as 2-3 individuals. Our estimates showed good convergent validity with directed faking results as well as significant between-culture variance.

Finally, our MGCFA demonstrates the replication of an MDPP inventory's higher-order factor structure across cultures. This provides a feasible method for using available dimension-level data to study measurement equivalence. Future research could expand our focus to invariance analysis using item-level data for non-adaptive MDPP tests, which would also be advantageous when data are not available for individual personality statements.

In conclusion, the present study demonstrates the importance of cross-cultural MDPP research as well as the challenges it presents. We hope that this demonstration will serve as a useful guide for researchers and inspire further developments in this domain. Given the largely untapped potential of MDPP testing for personnel selection, such developments could be invaluable.

References

Barrick, M. R., & Mount, M. K. (1991). The Big Five personality dimensions and job performance: A meta-analysis. *Personnel Psychology, 44*, 1-26.

Bollen, K. A. (1989). *Structural equations with latent variables.* New York: John Wiley.

Boyce, A. S., Conway, J. S. & Caputo, P. (2016). *Development and validation of Aon Hewitt's personality model and Adaptive Employee Personality Test (ADEPT-15).* New York: Aon Hewitt.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*, 233-255.

DeYoung, C. G., Quilty, L. C., & Peterson, J. B. (2007). Between facets and domains: 10 aspects of the Big Five. J*ournal of Personality and Social Psychology, 93*, 880-896.

Drasgow, F., Chernyshenko, O. S., & Stark, S. (2010). 75 years after Likert: Thurstone was right! I*ndustrial and Organizational Psychology, 3*, 465-476.

Drasgow, F., Stark, S., Chernyshenko, O. S., Nye, C. D., & Hulin, C. L. (2012). *Development of the Tailored Adaptive Personality Assessment System (TAPAS) to support Army selection and classification decisions* (Technical Report 1311). Fort Belvoir, Virginia: Army Research Institute for the Behavioral and Social Sciences.

Joseph, D. L., & Newman, D. A. (2010). Emotional intelligence: An integrative meta-analysis and cascading model. J*ournal of Applied Psychology, 95*, 54-75.

Lee, K., & Ashton, M. C. (2010). Psychometric properties of the HEXACO Personality Inventory. *Multivariate Behavioral Research*, 39, 329-358.

McDonald, R. P. (1989). An index of goodness-of-fit based on noncentrality. *Journal of Classification, 6*, 97-103.

Meade, A. W., Johnson, E. C., & Braddy, P. W. (2006, August). The utility of alternative fit indices in tests of measurement invariance. Paper presented at the annual Academy of Management conference, Atlanta, GA.

Roberts, J. S., Donoghue, J. R., & Laughlin, J. E. (2000). A general item response theory model for unfolding unidimensional polytomous responses. *Applied Psychological Measurement*, 24, 3-32.

Stark, S. (2002). A new IRT approach to test construction and scoring designed to reduce the effects of faking in personality assessment: The generalized graded unfolding model for multi-unidimensional paired comparison responses. Dissertation Abstracts International, 63, 1084.

Stark, S., Chernyshenko, O. S., & Drasgow, F. (2005). An IRT approach to constructing and scoring pairwise preference items involving stimuli on different dimensions: An application to the problem of faking in personality assessment. *Applied Psychological Measurement,* 29, 184-201.

Stark, S., Chernyshenko, O.S., Drasgow, F., & White, L.A. (2012). Adaptive testing with multidimensional pairwise preference items: Improving the efficiency of personality and other noncognitive assessments. *Organizational Research Methods*, 15, 463-487.

Steers, R. M., & Braunstein, D. N. (1976). A behaviorally-based measure of manifest needs in work settings. *Journal of Vocational Behavior,* 9, 251-166.

Steiger, J. H. (1989). EzPATH: *Causal modeling.* Evanston, IL: SYSTAT.

Tett, R. P., Jackson, D. N., & Rothstein, M. (1991). Personality measures as predictors of job performance: A meta-analytic review. *Personnel Psychology,* 44, 703-742.

VandeWalle, D. (1997). Development and validation of a work domain goal orientation instrument. *Educational and Psychological Measurement,* 57, 995-1015.

Table 1
Cultural Groups in the Social Desirability Sample

Country	Language	Sample Size
South Africa	Afrikaans	3
Lebanon	Arabic	4
Brazil	Portuguese	4
Czechoslovakia[a]	Czech	4
Australia	English	4
Canada	English	4
England	English	4
India	English	3
Indonesia	English	4
Philippines	English	5
Singapore	English	3
United States	English	6
France	French	5
Portugal	Portuguese	5
Germany	German	4
Hungary	Hungarian	7
Indonesia	Indonesian	6
Italy	Italian	5
Japan	Japanese	6
South Korea	Korean	5
India	Marathi	4
Malaysia	Malay	5
Poland	Polish	5
Romania	Romanian	4
Russia	Russian	3
Singapore	English	7
Spain	Spanish	6
Philippines	Tagalog	3
Thailand	Thai	3
Turkey	Turkish	4
Vietnam	Vietnamese	5

[a] *Combined sample from the Czech Republic and Slovakia*

Table 2
ADEPT-15 Styles, Dimensions, and Mappings to the Five-Factor Model

ADEPT-15 Style	ADEPT-15 Dimension	Five-Factor Model
Adaptation	Conceptual	Openness to Experience
	Flexibility	Openness to Experience
	Mastery	Unmapped to FFM
Task	Structure	Conscientiousness
	Drive	Conscientiousness
Interaction	Assertiveness	Extraversion
	Liveliness	Extraversion
Teamwork	Sensitivity	Agreeableness
	Cooperation	Agreeableness
	Humility	Unmapped to FFM
Emotional	Composure	Emotional Stability
	Positivity	Emotional Stability
	Awareness	Unmapped to FFM
Achievement	Ambition	Unmapped to FFM
	Power	Unmapped to FFM

Table 3
Results From the Social Desirability Study

ADEPT-15 Dimension	Total Items	ANOVA F-value (construct level)	ANCOVA F-value (construct level)	Significant ANOVA % (item level)
Ambition	102	1.718*	2.388**	34.3
Assertiveness	116	2.157**	4.630**	31.9
Awareness	66	1.834*	2.767**	33.3
Composure	98	1.830*	2.182**	60.2
Conceptual	81	1.701*	2.166**	40.7
Cooperativeness	113	2.692**	5.193**	40.7
Drive	84	2.592**	4.659**	58.3
Flexibility	98	2.172**	3.991**	34.7
Humility	101	1.533†	2.984**	39.6
Liveliness	100	1.293	1.965**	43.0
Mastery	85	2.094**	3.848**	35.3
Positivity	95	1.731*	2.467**	46.3
Power	112	1.548†	2.726**	31.3
Sensitivity	90	2.057**	4.050**	42.2
Structure	129	3.202**	4.905**	27.1
Overall	1470	2.839**	N/A	39.5

†$p < .10$
*$p < .05$
**$p < .01$

Table 4

Countries in the Measurement Invariance Sample

Country	Sample Size
Argentina	453
Australia	146
Brazil	919
China	7086
Columbia	449
Ecuador	278
India	10441
Malaysia	425
Mexico	2081
Thailand	789
United States	1284

Table 5

Results of the Multi-Group Confirmatory Factor Analysis

Invariance Model	χ^2	df	CFI	NCI	Gamma hat	$\Delta\chi^2$	ΔCFI	ΔNCI	ΔGamma hat
1. Configural	7046.739**	825	.827	.880	.967	–	–	–	–
2. Metric	7212.361**	915	.825	.879	.967	165.62**	.002	.001	.000
3. Scalar	12349.054**	1005	.684	.792	.942	5136.69**	.140**	.087**	.025**
4. Mean	14889.652**	1065	.615	.753	.930	2540.60**	.069**	.039**	.012**

Note. CFI = Comparative Fit Index; NCI = Non-Centrality Index. Δ = change from less constrained model.
**$p < .01$

Figure 1. Screenshot of MDPP items from ADEPT-15.
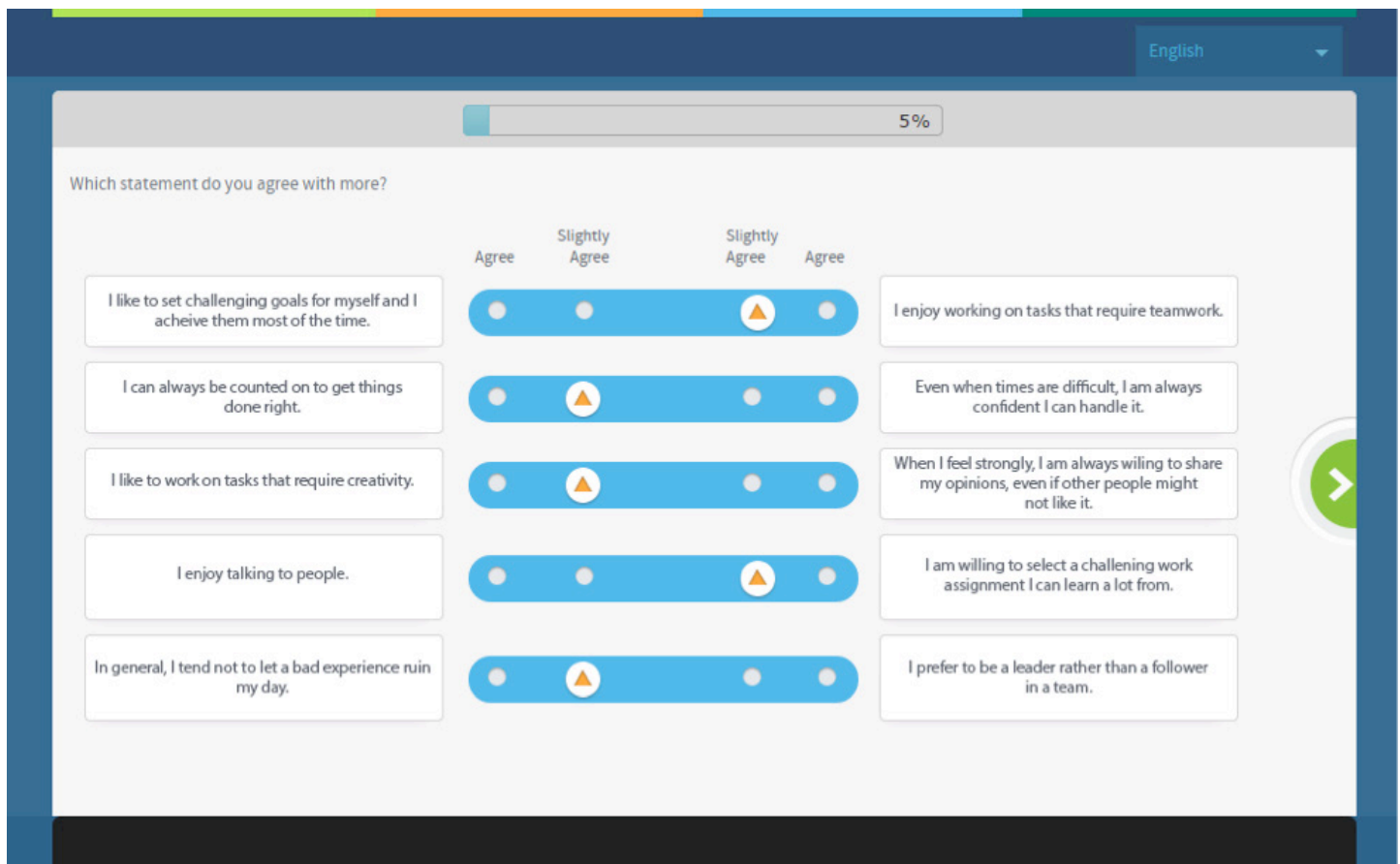
Personality and Motives: Investigating the Well-Being of Volunteers

Ashley J. Hoffman, The Chicago School of Professional Psychology
Drew Mallory, Katholieke Universiteit Leuven
Dr. Eleni V. Lobene, Aon Hewitt
Anthony S. Boyce, Aon Hewitt
Sarah A. Meeks, Wayne State University

Abstract: Volunteers are often used in the disbursement of humanitarian aid. However, little attention has been paid to the individual characteristics of people who choose to engage in this type of service. The current study seeks to explore the way personality and motives impact well-being of volunteers.

Humanitarian work psychology is a burgeoning sub-field of Industrial-Organizational Psychology. This sub-discipline is gaining traction, as researchers and practitioners alike are recognizing the dearth of research dedicated to the responsible and effective disbursement of humanitarian aid and development. As experts in understanding the impact of human resources, social scientists in particular are well-suited to address this gap in the literature. As such, the current study seeks to provide additional information regarding the personality and motive profiles of humanitarian aid workers, which may prove useful in the recruitment and selection, as well as training and development of such aid volunteers.

In comparison to many humanitarian aid topics, the general volunteerism literature tends to have a more robust body of publications. Recent literature has focused on volunteer recruitment (e.g., Boezeman & Ellemers, 2008; Peterson, 2004), medical aid workers (e.g., DeCamp, 2011; Green, Green, Scandlyn, & Kestler, 2009), cross-cultural considerations (e.g., Sherraden, Lough, & McBride, 2008), corporate social responsibility (e.g., Berens, van Riel, & van Rekom, 2007), and volunteer tourism, or "voluntourism" (e.g., Atkins, 2011; Barbieri, Santos, & Katsube, 2011). However, little has been done to outline the very basic individual characteristics that indicate a likelihood that someone might volunteer initially. Specifically, there is a dearth of research devoted to investigating the personality characteristics, motives, and overall life satisfaction of volunteers. As it is likely that volunteering will continue long into the future, it is important that social scientists begin to understand the individual differences inherent to those volunteers in order to ensure the effectiveness of such aid, and the well-being of individuals engaging in this work.

Computer adaptive personality testing. An additional aspect of the current study that is particularly compelling is the use of a newly developed personality inventory called the ADEPT-15. ADEPT-15 utilizes a computer-adaptive approach to personality testing. This previously underexplored approach uses the basic principles of computer adaptive testing to create a unique personality profile that is presumed to be particularly specific to the participant. Computer adaptive testing (CAT) has become a well-established approach to investigating cognitive ability in many organizations (e.g. Aon Hewitt, IBM, Proctor & Gamble) due to its unique ability to provide detailed outcomes of participants' capabilities. CAT relies heavily on the theoretical underpinning of Item Response Theory (IRT), which is a statistical approach that seeks to identify the pure ability of participants by using technology to present participants with items drawn from answers to questions answered previously in the survey. CAT is advantageous because of the potential for decreased test length and increased measurement accuracy (e.g., Hambleton, Swaminathan, & Rogers, 1991; Wainer, 2000).

To develop ADEPT-15, researchers pulled traits from existing Five-Factor Model personality measures (Hogan Personality Inventory, Hogan & Hogan, 2007; Occupational Personality Questionnaire (OPQ32), Brown & Bartram, 2009; 16PF, Conn and Rieke (1994); Facet-5, Buckley & Williams, 2011; International Personality Item Pool, Goldberg et. al, 2006; NEO-PI, Costa & McCrae, 2005). Additionally, aspects of personality were included that are considered important to job performance and leadership roles, but that were not neatly captured by the aforementioned five-factor tools, including ambition and power. From perusal of the five-factors models and the

additional variable consideration, researchers identified fifteen aspects of personality believed to be of relevance. These fifteen aspects were then grouped into six higher-order styles. These styles are shown in Figure 1. After identifying these traits and styles, researchers compiled thousands of potential assessment items. These items are then presented in unique item strings, individually tailored to each participant based on previous item response in order to hone in on specific personality traits of individuals. Of specific interest to researchers is the adaptation style of teamwork. This style is comprised of humility, cooperation, and sensitivity. The current study also includes the aspect of positivity, as it seems likely that this aspect should impact an individual's overall subjective well-being. Definitions of each aspect are provided in Table 1.

Subjective well-being. Distinct from a positive or negative affect, subjective well-being (SWB) has been proposed as a proxy for both affective and cognitive aspects of happiness, and in most cases is considered an overall cognitive judgment of one's satisfaction with his or her life as a whole (e.g. Steel, Schmidt, & Shultz, 2008). The current study approaches SWB from a cognitive approach, whereby individuals make a global assessment of their life, and in particularly, the satisfactory nature of their life as a whole. Previous research has indicated a significant relationship between cognitive SWB and personality factors, particularly using the personality traits of extraversion and emotional stability as predictors (Schimmack, Schupp, & Wagner, 2008).

Motives for volunteering. Each volunteer may experience measurably different underlying attitudes that induce the action of volunteering. For example, volunteers' motives could be related to gaining experience relevant to their occupation, or to the fact that the opportunity aligns with their personal values. Researchers in the domain of volunteering have used the functional approach to motives, whereby different people engage in the same behaviors for a variety of different reasons, and based on differing attitudes and values (Snyder, 1993). Clary and colleagues (1998) have examined how best to quantify the motives of volunteers. Specifically, they developed a measure of volunteer motives called the Volunteer Functions Inventory (VFI.) The VFI was developed to assess certain factors related to volunteer motives, and each motive is characterized by a unique way of thinking. Of particular interest in the current study are volunteers motivated by personal values. These volunteers tend to express altruistic intentions and a great deal of concern for others. As such, it seems that volunteers with personality traits related to compassion and positivity, who are also motivated by strong values, should experience more satisfaction with their lives because they are able to have those value motives fulfilled. Therefore, volunteer motives will be assessed in the current study using the VFI in order to investigate the following research questions:

Research questions 1a-d: Does value motives will mediate the relationship between personality styles (1a: co-operation; 1b: sensitivity; 1c: humility; 1d: positivity) and subjective well-being?

Method
Participants
Researchers used the online participant tool Mechanical Turk in order to collect data regarding the individual characteristics of people who had volunteered at some point in the last six months. After removing participants who did not spend adequate time to complete the survey, the sample included a total of 253 respondents. These individuals were compensated for their participation upon completion of the study, and data was collected from volunteers globally. The majority (63.2%) of participants were Caucasian while 20.9% were Asian or Pacific Islander, 7.1% were African American, 4.7% were Hispanic, 3.6% were American Indian, and 0.4% Other. Most respondents had a 4-year College Degree (43.9%), while roughly 16% had a High School Diploma, 2-year College Degree, or Master's Degree. Four percent or less reported having a Professional Degree, Some High School, or a GED. There was a fairly even split between male (50.6%) and female (49.4%) participants. Finally, the majority of participants were from the United States (81.8%) while 16.2% were from India and less than a percent were from Macedonia, Nepal, Canada, and Sri Lanka.

Measures
ADEPT-15.
ADEPT-15 is a computer adaptive personality test. This assessment draws from a bank of thousands of possible items in order to hone in on the most accurate description of a participant's personality. As such, each survey will be unique to the participant based on responses to previous items in the survey.

Subjective Well-Being. Subjective well-being was measured using a short scale investigating overall happiness and satisfaction with life (Diener, E. D., Emmons, R. A., Larsen, R. J., & Griffin, S., 1985).

Volunteer Functions Inventory (VFI). Volunteer motives were assessed using a 30 item measure (Clary, et al., 1998). This measure is divided into 6 sub-sections, which include protective, value-based, career, social, understanding, and enhancement motives.

Results

Table 2 displays the means, standard deviations, and reliabilities (test-retest or coefficient alphas) for all of the variables included in our study. Each measure achieved an acceptable level of reliability (Nunnally & Bernstein, 1994). Given the unique multidimensional forced choice adaptive format, it is most appropriate to assess the reliability of ADEPT-15 using test-retest reliability as opposed to other traditional classical test theory estimates of reliability (e.g., Cronbach's alpha, split-half reliability). The interested reader should refer to Boyce, Conway, & Caputo, 2014 for additional details on the psychometric properties of the assessment. As our review of the literature indicated, value-based motives for volunteering related positively to each of the personality variables included, in addition to wellbeing. Of the personality scales, cooperative personality and humble personality did not significantly correlate with wellbeing, whereas sensitive and positive personalities showed weak to moderate correlations respectively.

To test for mediation, we used Hayes' (2012) PROCESS macro (Model 4) for SPSS (v. 22), which can conduct simple mediation analyses using ordinary least squares path analysis. Following the advice of Hayes (2012) we obtained bias-corrected bootstrapped confidence intervals for the indirect effects (using 5,000 bootstrap samples). Participants with missing values for variables of any analyses were automatically excluded from all analyses by PROCESS (n=1), which reduced the sample size to 252. Analyses were all two-tailed with alpha set at .05. As shown in Table 3, three of these analyses were upheld: cooperation, sensitivity, and humility indirectly influenced volunteers' subjective well-being through the effect of value-driven motives to volunteer. With the addition of values motive, positivity no longer showed a significant relationship with well-being. Respondents higher in cooperative (B = .77), sensitive (B = 2.06), and humble tendencies (B = .93) were more likely to be motivated to volunteer due to their own personal values, and those who were motivated to volunteer due to their values were more likely to experience higher wellbeing (B = .43; .36, .48, respectively).

Table 4 displays the bias-corrected bootstrap confidence intervals based on 5,000 bootstrap samples computed for the indirect effects (B = .33, .74, .45) were entirely above zero for each of these predictors. For cooperative and sensitive personalities, there was no evidence personality influenced wellbeing independent of its effect on values (cooperative: B = 0.20, p = .84; sensitive B = 1.23, p = .20). However, when taking value-based motives for volunteering into account, the main effect remained significant, with those scoring higher on the humble personality showing significant negative relationship to wellbeing (B = -2.28, p = .03). The latter was somewhat surprising as it seems that people with high humility scores should experience more subjective well-being. Nonetheless, when motives for volunteering are value-based, this effect is diminished, suggesting that for individuals with more humble personalities, having a value-based motive for volunteering may act as a suppressor of a potentially negative effect on wellbeing.

Discussion

The current study sought to determine specific personality indicators that may be relevant to the overall well-being of those that volunteer, particularly with the consideration of one's motive for volunteering. It was determined that three of our research questions were supported, such that respondents who scored higher on cooperativeness and sensitivity were more likely to be motivated by their own values, and experience a greater sense of overall well-being. Cooperativeness and sensitivity are traits related to teamwork, and personal values are marked by altruistic intention and concern for others. It logically follows suit that volunteers with strengths related to working with and understanding others, who are able to have value motives related to serving others fulfilled, would experience greater satisfaction with their lives. Such individuals who experience alignment between themselves and their circumstances are fortunate and reap psychological benefits.

In addition, those people who scored highly on humility scales experienced a negative relationship with well-being when their motives were value driven. This finding may be related to the fact that humble people may be less effective at advocating for their own interests and could experience frustration when they are motivated based on their values, but unable to effectively pursue them. Since volunteer work often requires engagement in unfamiliar situations, those people who are able to be bold and assertive could fair better than those who feel uncertain or meek.

There were some limitations to consider in this study. First, researchers collected their data using the online site Mechanical Turk. There has been some conflicting opinion about the use of MTurk as a data collection tool, and the researchers are aware of this potential limitation. However, due to the computer adaptive nature of the ADEPT-15 measure, along with its built in social desirability controls, it is our hope that this limitation is offset. An additional limitation is the nature of responding to items regarding well-being. There is again the opportunity for participants to respond in a way that overestimates the global cognitive well-being they are experiencing. This potential for "faking good" could impact the outcomes we have found.

Despite these limitations, it seems to be that those volunteers who are engaging in volunteer opportunities due to their own strong sense of morals and values, and who have a great deal of cooperation and sensitivity, tend to experience an overall sense of higher subjective well-being. This information can begin to provide a clearer picture of the individual characteristics of volunteers, and future research should consider more aspects of personality and motives in order to further understand the types of people who engage in volunteer pursuits.

References

Atkins, S. G. (2011). Smartening-up voluntourism: SmartAid's expansion of the Personality-focused Performance Requirements Form (PPRF). *International Journal of Tourism Research.* 14(4): 369-390.

Barbieri, C., Santos, C.A., & Katsube, Y. (2011). Volunteer tourism: Oh-the-ground observations from Rwanda. *Tourism Management*, 33: 509-516.

Berens, G., van Riel, C.B.M., & van Rekom, J. (2007). The CSR-quality trade-off: When can corporate social responsibility and corporate ability compensate each other? *Journal of Business Ethics.* 74: 233–252.

Boezeman, E. J., & Ellemers, N. (2008). Pride and respect *in* volunteers' organizational commitment. *European Journal of Social Psychology*, 38: 159-172.

Boyce, A.S., Conway, J.S. & Caputo, P. (2014). Development and validation of Aon Hewitt's personality model and Adaptive Employee Personality Test (ADEPT-15). New York: Aon Hewitt.

Brown, A., & Bartram, D. (2009). Development and Psychometric Properties of OPQ32r. Retrieved from http://www.shlsolutionpartner.com/au/resources/NEWOPQ32TechManualsupplement.pdf.

Buckley, N. L., & Williams, R. J. (2011). Facet5 Technical Summary. Retrieved from http://www.facet5global.com/_literature_102975/Facet5_Technical_summary.

Clary, E. G., Snyder, M., Ridge, R. D., Copeland, J., Stukas, A. A., Haugen, J., & Miene, P. (1998). Understanding and assessing the motivations of volunteers: a functional approach. *Journal of personality and social psychology*,74(6): 1516.

Conn, S. R., & Rieke, M. L. (1994). 16PF Fifth Edition Technical Manual. Champaign, IL: Institute for Personality and Ability Testing.

Costa, P., & McCrae, R. (2005). The NEO-PI-3: a more readable revised NEO P*ersonality Inventory. Journal of Personality Assessments*, 84(3): 261-270.

DeCamp, M. (2011). Ethical review of global short-term medical volunteerism. *HEC Forum.* 23: 91–103.

Diener, E. D., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The satisfaction with life scale. *Journal of personality assessment*, 49(1): 71-75.

Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. C. (2006). The International Personality Item Pool and the future of public-domain personality measures. *Journal of Research in Personality,* 40: 84-96.

Grönlund, H., Holmes, K., Kang, C., Cnaan, R., Handy, F., Brudney, J., et al. (2011). Cultural values and volunteering: A cross-cultural comparison of students motivation to volunteer in 13 countries. *Journal of Academic Ethics*, 9(2): 87–106.

Hambleton, R.K., Swaminathan, H., & Rogers, H.J. (1991). *Fundamentals of item response theory.* Newbury Park, CA: Sage.

Hayes, A. F. (2012). PROCESS: A versatile computational tool for observed variable mediation, moderation, and conditional process modeling [White paper]. Retrieved from http://www.afhayes.com/public/process2012.pdf

Hogan, J., Barrett, P., & Hogan, R. (2007). Personality measurement, faking, and employment selection. *Journal of Applied Psychology*, 92, 1270-1285.

Hustinx, L. (2010). Weakening organizational ties? A classification of styles of volunteering in the Flemish Red Cross. *Social Service Review,* 79(4), 624-652.

Nunnally, J. C., & Bernstein, I. H. (1994). Psychometric Theory, 226.

Peterson, D.K. (2004). Recruitment strategies for encouraging participation in corporate volunteer programs. *Journal of Business Ethics*. 49: 371–386.

Schimmack, U., Schupp, J., Wagner, G.G. (2008) The influence of environment and personality on the affective and cognitive component of subjective well-being. *Social Indicators Research*, 89: 41–60.

Sherraden, M.S., Lough, B., & McBride, A.M. (2008). Effects of international volunteering and service: Individual and institutional predictors. *Voluntas*. 19: 395–421.

Snyder, M. (1993). Basic research and practical problems: The promise of a "functional" personality and social psychology. *Personality and Social Psychology Bulletin*, 19: 251–264.

Steel, P., Schmidt, J., Shultz, J. (2008). Refining the relationship between personality and subjective well-being. *Psychological Bulletin*, 134(1): 138–161

Wainer, H. (2000). *Computerized adaptive testing: A primer.* Mahwah, New Jersey: Lawrence Erlbaum Associates.
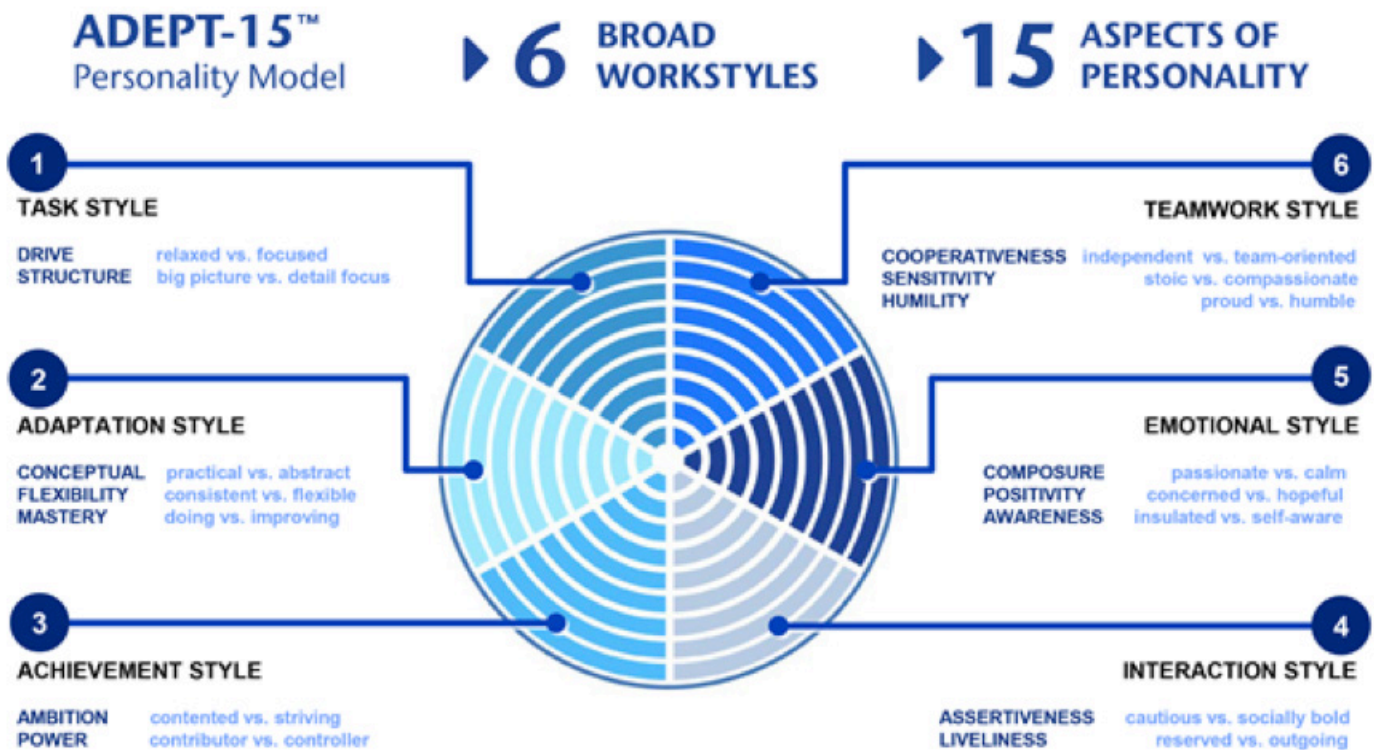
Figure 1
ADEPT-15 styles

Table 1

ADEPT-15 style aspect	Definition
Humility	This aspect of personality measures the extent to which someone is modest and genuine. High scorers tend to be humble and unselfish, though may be less effective in advocating for own interests. Low scorers are proud, cunning, and can be manipulative, but are also bold and can be adept at managing situations requiring tact and posturing.
Cooperation	This personality aspect reflects the extent to which someone is cooperative and trusting. People who score high tend to be team oriented and accommodating, but can be sometimes be socially naïve. Those who score low tend to be more independent-minded and less interested in teamwork, but also less likely to be taken advantage of by others.
Sensitivity	This personality aspect reflects the extent to which someone is compassionate, caring, and understanding. Those who score high tend to be warmhearted, patient, and tolerant, but may have difficulty providing negative feedback or being firm with others. Those who score low tend to be stoic and tough-minded, but can be frank and direct.
Positivity	This personality aspect reflects the extent to which someone is happy, optimistic, and resilient. High scorers tend to be hopeful and positive, but often ignore roadblocks. Low scorers can be pessimistic and overwhelmed with obstacles, but tend to be more realistic.

Table 2

Means, Standard Deviations, Correlations, and Internal Consistency Reliabilities of Study Variables

Variable	M	SD	1	2	3	4	5	6
1. Cooperation	.08	45	(0.66)					
2. Sensitivity	.17	.49	.29**	(0.66)				
3. Humility	.08	.43	.01	.24**	(0.57)			
4. Positivity	.14	.50	.17**	.17**	.02	(0.70)		
5. Well-being	24.40	7.01	.04	.14*	-.11	.43**	(0.92)	
6. Values motive	21.30	2.87	.12*	.35**	.14*	.17**	.18**	(0.83)

N = 252.

*$p < .05$. ** $p < .01$.*

Note: Established test-retest reliability estimates are included for Cooperation, Sensitivity, Humility, and Positivity. Internal-consistency reliability estimates (coefficient alphas) are included for Well-being and Values motive from this sample. These are all reported in the parentheses along the diagonal.

Table 3
Regression Results

Predictor	Value-driven Motive			Subjective Wellbeing		
	B	*SE*	R²	*B*	*SE*	R²
Cooperation	.77**	.40	.01*	.20	.97	.00
Sensitivity	2.06**	.34	.13**	1.23	.94	.04**
Humility	.93*	.42	.02*	-2.28*	1.03	.05**
Positivity	.96*	.36	.03*	5.76**	.82	.19**

* $p < .05$. ** $p < .01$.   a

Table 4
Estimates and Bias-Corrected Bootstrapped 95% Confidence Intervals for the Indirect Effect of Value-driven Motive for Volunteering on Volunteer Wellbeing

Predictor	Indirect Effect on Wellbeing	
	Estimate (SE) [a]	*CI*
Cooperation	.33 (.20)	[.03, .63]
Sensitivity	.74 (.34)	[.14, 1.52]
Humility	.45 (.25)	[.08, 1.07]
Positivity	.26 (.18)	[.00, .72]

Bootstrapped estimates for the standard error *(SE)* are presented.

A New IPAC Regional Chapter:
The Southwest Assessment Group

Erin Smith, Test Validation Analyst, City of Dallas

Greetings IPAC members! During this year's annual conference, a group of members met to discuss the idea of forming an IPAC chapter for Texas, Louisiana, Arkansas and Oklahoma. Our group has tentatively been named The Southwest Assessment Group (SWAG). Our goal is to create a regional network of personnel assessment professionals in this region of the country. As Vice-Chair of this new chapter in the IPAC family, I am proud to share who we are, what we plan to do, and how you can get involved as an IPAC member.

The creation of SWAG was inspired by the success that our friends at GLEAN (Great Lakes Employment Assessment Network) have experienced in providing training and resources for their members in the Great Lakes region. GLEAN was established in 2013 as IPAC's first regional group. Since then, GLEAN has connected hundreds of human resources, civil service, and assessment professionals through their biannual training events.

SWAG hopes to offer professionals in our region the same opportunities that GLEAN offers its members; however, we will tailor our efforts to meet the unique challenges that professionals in our region often encounter. If you are interested in learning more about our chapter and live in or near the southwest region (and trust me, we don't mind stretching the definition of "southwest" a bit), we welcome you to contact our chair, Andrew Yurkon, (andrew.yurkon@dallascityhall.com) or me (erin.smith@dallascityhall.com) to learn more about our chapter and how you can get involved. We will keep the IPAC community updated with our events and activities. We look forward to our first steps in expanding the IPAC family.

From the Perspective of a New Hire: Onboarding Options and Opportunities

Kelsey Stephens, Manager, People Science & Optimization, Macy's Inc.

Onboarding helps new hires adjust to the cultural, social, and performance aspects of their jobs so they can quickly become productive, contributing members of their organization. New employees are both eager and expecting to learn about their new company; therefore, it certainly makes good business sense to capitalize upon this time in an employee lifecycle by strategically leveraging onboarding programs. Research and conventional wisdom both suggest that employees are working with approximately 90 days to prove themselves in a new job. The faster new hires feel welcomed and prepared for their jobs, the faster they will be able to successfully contribute to the organization's mission.

This past spring, I became a new member of the People Science and Optimization team at Macy's Inc. Not only did my newest colleagues greet me with a warm welcome, including a visit to the Red's Opening Day Parade and a lunchtime visit to famous Skyline Chili, there was no denying that the team was well prepared for my first day on the job.

After taking care of the basic HR paperwork, my supervisor handed me an onboarding report. This report was explicitly designed to help me successfully assimilate into the team. The content in the report was generated by a dozen stakeholders, partners, and peers, who had completed a short survey. The survey included nine optional questions to which respondents were encouraged to respond openly and candidly; the responses told me not only why the team was excited to have me onboard—but also why they may be nervous to have me join the team! The report stated that this integration plan was created in order to shorten the time it would take me to make a strong impact.

I also was armed with the inside scoop as it relates to the expectations my new team members had for my first 30 to 90 days with Macy's. The responses told me what challenges lie ahead of me in terms of relationship building and assimilation progress. The survey even prompted respondents to respond to the elephant in the room — it was so helpful to hear about the significant threats to the business firsthand from my new colleagues.

Soon, I learned that I received this report because I was part of a pilot group. This is a smart strategy, as rolling out to the larger organization could present challenges without first testing the initiative. However, breaking it down into a manageable, targeted group allowed for a test-and-learn design. Needless to say, I have provided nothing but accolades to the team and have encouraged a larger roll-out of the onboarding report. In total, the report helped me to better understand the cultural, social, and performance factors that made up my new position.

The Professional and Scientific Affairs Committee

Dennis Doverspike, Chair

In the June edition of the ACN, the role of the Professional and Scientific Affairs Committee (PSAC) was discussed and feedback was requested from IPAC members. No responses were received. I am beginning this column by once again requesting suggestions, comments or any other feedback you might have.

Creating IPAC Podcasts

The PSAC would like to start offering **podcasts**. For those who aren't familiar with podcasts, podcasts are similar to radio shows, and may include video capture in addition to audio. Some ways in which you might assist with the podcasts are:
• Suggest topics
• Create or assisting in creating a podcast
• Share an existing podcast or webinar

We are specifically interested in hearing from graduate students or early-career professionals who may have ideas for a podcast. For example, you might want to interview your faculty, assessment practitioners or consultants for their perspectives on past or contemporary assessment related issues.

IPAC Monographs

We are also interested in continuing the IPAC **monograph** series. Monographs are often longer than the regular research articles and typically focus on a single specialized question or area often summarizing a large body of scientific research. IPAC monographs are usually written in a style that is understandable to assessment professionals at all levels of expertise. Example of past monographs can be found at the IPAC Resources Page: http://www.ipacweb.org/library.

Updating existing IPAC monographs would also be helpful. Many of the IPAC monographs are somewhat dated. Many of the monographs are still relevant but could use some revisions and updates.

Moving Forward

The PSAC would like to create some type of synergy and synchronization between this column and the podcasts and webinars. There are a lot of assessment related topics opinions, that may be of interest to IPAC Members. We could start with what many would consider the basics of what we do – the Merit Principle and the Assessment of Merit.

Personnel Assessment and Decisions (PAD)

We would like to invite you this opportunity to direct your attention to two special issues of our journal, PAD. The topics will be:
• Applications of Judgment and Decision Making to Problems in Personnel Assessment; and
• Advanced Technologies for Personnel Assessment.

You can find submission guidelines for the special issues at https://scholarworks.bgsu.edu/pad/.

We would like to see more practitioners and others with applied assessment experience serve as reviewers for PAD. Serving as a reviewer is an easy way to contribute to IPAC and make a professional contribution.

To provide suggestions for future topics for this column, volunteer to create or assist with podcasts, monographs or review for PAD, please contact me at dennisdoverspike@gmail.com.

If you prefer using social networking websites, you may contact me at:
LinkedIn: https://www.linkedin.com/in/drdennisdoverspike/
Google Plus: https://plus.google.com/u/0/+DennisDoverspike
Twitter: https://twitter.com/ddoverspike
Facebook: https://www.facebook.com/dennis.doverspike

Legal Watch

Ryan O'Leary
PDRI a SHL Company

SCOTUS Watch

The U.S. Supreme Court issued several decisions in the term that ended June 2018 that are relevant to employment law.

Class Action Cases

China Agritech, Inc. v. Resh. China Agritech manufactures and sells farming products. It began listing shares on the NASDAQ in 2005 and in 2011 issued a public financial filing that resulted in several lawsuits. A market research company reported that the filing made grossly inflated claims of China Agritech's revenue and value. Shareholders sued for having been misled and China Agritech faced a series of class actions. The third was by filed Resh (after the first two) and has relevance for employment law. The issue in this case was whether Resh could take advantage of a tolling rule, a rule allowing more time to file a case. If not, the Resh case would have been filed too late and would have to be dismissed. In a unanimous decision, the Court ruled that the filing of a putative class action (a lawsuit brought by one or more named plaintiffs on behalf of a potential group of similarly situated individuals [known as a class] who allegedly suffered a common claim) does not toll (i.e., pause) the statute of limitations for follow-on class action when class certification is denied. In other words, this means that plaintiffs may not resurrect a failed class action by filing another class action after the limitation period has expired.

In this decision, the Court distinguished its precedent from two earlier cases (*American Pipe and Construction Co. v. Utah, 1974; Crown, Cork & Seal Co. v. Parker,* 1983) which held that the filling of a class action does toll the limitations period for individuals who are seeking to intervene in the suit or to file their own individual claims after class certification is denied. Legal analysis suggests that these two rulings were meant to encourage class action cases. Class action cases generally help the courts save time and money. As such, courts would rather hear one big case on the same issue than many, smaller individual cases on the same issue. If there is a class action going on that would include an individual's claim but the individual is considering otherwise filing an individual lawsuit, the individual can wait to see how the class action turns out – the time period during which one must file is paused (i.e., tolled) while the class action is ongoing. Court efficiency appears key and class actions which save the court resources should go first. Individuals are encouraged to see what happens in the class action case before pursuing their individual claims. Class claims should be ruled earlier and the Court will not allow class claims to be filed later.

As it relates to employment law, in their opinion the Court noted that following the denial of nationwide class certification in Wal-Mart Stores, Inc. v. Dukes (2011) (the class of approximately 1.6 million women who claimed gender discrimination in pay and promotions at Wal-Mart which was denied certification because they did not have enough in common), numerous plaintiffs had either amended the original complaint to repeal subclasses or separately asserted geographically regional subclasses within the limitation period. The *China Agritech, Inc. v. Resh* suggests we may see future plaintiffs file multiple parallel complaints raising different possible putative subclasses with employers moving to stay the subclass actions until the court resolves whether to certify the larger class.

Epic Systems Corp. v. Lewis. In May 2018, the Court ruled that companies can use arbitration clauses in employment contracts to prohibit workers from banding together and taking legal action over workplace issues. This means that employees who sign arbitration agreements can be precluded from participating in class action lawsuits and must therefore litigate their cases on individual bases. The 5 to 4 vote upheld the use of arbitration agreements in the workplace. The court majority stated that the ruling was a logical reading of the law and reflected Congress' preference for using arbitration to avoid costly and time-consuming litigation. The Court had earlier ruled that companies doing business with consumers may require arbitration and forbid class actions in their contracts. Arbitration clauses with class action waivers are now commonplace in contracts for things like cellphones, credit cards, and rental cars. The issue in Epic Systems Corp v. Lewis was whether these same principles apply to

employment contracts.

The decision applies to three separate cases relating to three employers: Epic Systems, Ernst and Young, and Murphy Oil. In all three cases, plaintiffs filed in federal court stating that their employers violated the Fair Labor Standards Act by not paying them overtime. However, in each case employees all signed arbitration agreements stating that they would arbitrate any disputes before a single arbitrator whose decision would be final and binding. Claims relating to different employees had to be heard in separate hearings.

Writing for the majority, Justice Gorsuch stated that in the Federal Arbitration Act (FAA) Congress has instructed the federal courts to enforce arbitration agreements according to their terms – including terms providing for individual proceedings. Plaintiffs had asserted that the National Labor Relations Act (NLRA) makes illegal any contract that denies employees the right to engage in "concerted activities" for the purpose of "mutual aid and protection" and therefore some form of collective action cannot be prohibited. The Court asserted that since the NLRA does not expressly approve or disapprove of arbitration, the FAA prevails. Some legal scholars predict the impact of the Court's ruling will be to largely eliminate the threat of employee class actions in cases such as a failure to pay overtime or systemic discrimination.

Hiring of Federal Administrative Law Judges

In Lucia v. Securities and Exchange Commission, the Court ruled in a case involving federal Administrative Law Judges (ALJs) at the agency. ALJs conduct trial-like hearings within federal agencies related to disputes over decisions such as claims for benefits and enforcement actions against individuals or businesses. This case was brought by a former financial advisor, Raymond Lucia, who promoted a retirement strategy he called "Buckets of Money" through radio shows, books, and seminars. The strategy suggested that retirement investors should first sell safer investments, giving riskier investments time to grow. In 2012 the SEC charged Lucia with violating federal law and SEC rules, claiming he mislead investors in presentations to potential clients. Lucia and his company were fined $300,000 and he was barred from working as an investment advisor. Lucia challenged the case and argued that the ALJ who heard his case was improperly appointed.

Central to the case is the Appointments Clause in Article II, Section 2 of the Constitution which states in part "[The President] shall nominate, and by and with the Advice and Consent of the Senate, shall appoint…Officers of the United States, whose Appointments are not herein otherwise provided for, and which shall be established by Law: but the Congress may by Law vest the Appointments of such inferior Officers as they think proper, in the President alone, in the Courts of Law, or in the Heads of Departments." At issue is whether SEC ALJ's are "officers" and in particular "inferior officers" who may be appointed by heads of departments. SEC ALJ's were selected by the Chief Judge and approved by the Commission's personnel office as opposed to being appointed by the Commissioners.

The Court of Appeals for the D.C. Circuit held that SEC ALJ's are not "Officers of the United States" but are instead mere employees, officials with lesser responsibilities who are not subject to the Appointments Clause. However, the Department of Justice who had long contended that the judges were employees and not officers, switched positions and urged the Court to grant review in the case even though it had won in the appeals court.

In a 7 to 2 decision the Supreme Court ruled that SEC ALJ's are officers rather than mere employees since the judges exercise significant authority in hearing and ruling on disputes. It did not matter to the Court that the judges' decisions were subject to review by the Commission. Since the ALJ's were appointed by staff members rather than by the Commissioners, their selection violates the Constitution's Appointment Clause which requires "inferior officers" to be appointed by the president, the courts, or heads of departments. The Commission itself is a "head of department", while its staff members are not. Since the SEC ALJs were not properly appointed, Mr. Lucia was entitled to a new hearing. This decision has a significant impact on how ALJs are selected and appointed.

Union Dues

In *Janus v. AFSCME Council 31*, Mark Janus (a child-support specialist at the Illinois Department of Health and Family Services) sued the American Federation of State, County, and Municipal Employees (AFSCME) union. Janis contended that he did not agree with AFSCME's positions and should not be forced to pay fees to support

its work. Under Illinois law, state employees represented in a bargaining unit are not compelled to be members of the union or pay union dues. However, they must pay an "agency fee", an amount equal to that portion of union member dues spent directly on bargaining and administration of the bargaining agreement.

The Court ruled in a 5 to 4 decision that public sector employees who are non-members of a union cannot be legally required to pay an agency or "fair share" fees as a condition of employment. This decision overturned a 40-year-old precedent established in the *Abood v. Detroit Board of Education* decision that said that states could allow public-employee unions to collect fees from non-members to cover the costs of workplace negotiations over salaries and benefits but not the union's political activities.

Racial Balancing and Affirmative Action in Academic Admissions

In *Students for Fair Admissions Inc. (SFFA) v. Harvard,* SFFA (which includes more than a dozen Asian-American students who applied to Harvard and were rejected) has accused Harvard of intentionally discriminating against Asian-American applicants by limiting their admission numbers each year. The suit, initially filed in Federal District Court in 2014, accuses Harvard of "racial balancing" – keeping roughly the same distribution of racial groups year after year despite changes in application rates and qualifications. Harvard denies that it conducts racial balancing or discriminates against Asian-Americans. They claim they use a "whole person evaluation" and that race is one of many factors considered in the pursuit of diversity.

Multiple times the U.S. Supreme Court has affirmed that universities may take race into account as one factor among many to achieve a diverse class. But there are limits on what colleges may do. The Court prohibits racial quotas and encourages colleges to consider whether they can achieve their goals through race-neutral alternatives such as using financial aid and other recruiting tools to ensure socioeconomic and geographical balances.

SFFA is viewed by many as an anti-affirmative–action group and the lawsuit part of an ongoing effort to do away with race-conscious affirmative action. In August 2018, the Department of Justice filed a legal brief in the case lending its support to the plaintiffs. Alternatively, a large number of Harvard supporters have filed briefs in the case, claiming that a failure to consider race would effectively threaten diversity at all American colleges.

The case is set to go to trial in October and may have far-reaching implications for the nation's colleges and universities that consider race in their admission processes. The case may end up at the Supreme Court, which is likely to be more conservative than in 2016, when it upheld narrowly tailored race-conscious admissions in *Fisher v. University of Texas at Austin* in a 4-3 decision.

Sexual Orientation as a Protected Class Under Title VII

A case that the Society for Human Resource Management (SHRM) cited as one of the top 10 employment cases of 2017 has finally come to an end. Kimberly Hively, the professor in the Seventh Circuit's landmark ruling that protection under the Civil Rights Act extends to sexual orientation has settled with her former employer, Ivey Tech Community College. The parties filed a joint mediation summary on August 1, 2018 announcing the settlement. The terms of the agreement were not released.

Kimberly Hively was an openly lesbian, part-time adjunct professor at Ivy Tech Community College. In 2014 she claimed that she had been repeatedly denied full-time employment and promotions because of her sexual orientation in violation of Title VII of the Civil Rights Act of 1964. The District Court dismissed her case, ruling that Title VII does not recognize sexual orientation as a protected class. The case then went to the 7th Circuit Court of Appeals where a three judge panel affirmed the District Court's ruling. Hively filed for a rehearing and the majority of the 7th Circuit found that the Civil Rights Act protection does prohibit discrimination based on sexual orientation and became the first U.S. Court of Appeals to rule that sexual orientation discrimination was prohibited under Title VII.

Improving Physical Test Validation from a Cognitive Viewpoint

Philip T. Walmsley & Randolph K. Park
U.S. Customs and Border Protection

Author Note: The views expressed in this paper are those of the authors and do not necessarily reflect the views of U.S. Customs and Border Protection or the U.S. Federal Government. Correspondence should be addressed to Phil Walmsley at walmsley.phil@gmail.com.

Abstract

Job analysis is foundational for establishing valid use of selection instruments. Physical test validation uses task statements requiring greater detail than task statements typically used to support cognitive assessments. Physical task statements are situation-bound to elicit reliable judgments based on a clear understanding of the actions and consequences of the actions. In this column, we rethink procedures ordinarily applied in the cognitive domain to inform validation work in the physical domain.

Introduction and Organizing Theme

In this column, we examine procedures for validating physical ability employment tests through the lens of common practices applied in support of cognitive and psychological employment tests. Professionals working in personnel assessment are familiar with procedures for evaluating assessment tools that measure cognitive and personality-based attributes (for shorthand, we refer to these as the psychological domain). The same validation procedures have been adapted to support measurement of attributes in the physical domain (e.g., muscular strength and endurance, aerobic capacity, dynamic movement).

Yet, application of validation methods to physical test scores uncovers at least several issues that we believe are not well recognized in applied research. Our objective is to highlight several issues within physical job analysis that pertain to test validation with the purpose of influencing thoughtful future practice. We have found that the exercise of comparing approaches used in cognitive and physical domains provides a useful mechanism for analyzing decisions that must be made during validation studies.

Many occupations in the global economy, such as those in labor and skilled trades, public safety, and the transportation industry, continue to require job incumbents to use physical attributes in pursuit of safe performance. Our perspective is informed primarily through work with law enforcement and military professionals. We focus this column on considerations involving job analysis as forming the foundation for a validity study.

Specificity in Job Analysis

Determining the appropriate specificity for job descriptors (e.g., tasks, KSAOs, competencies) is a central issue in any job analysis. Analysts must make choices about what information should be gathered and how it will be used, and these choices should be driven by the purpose for how the data will be used. For assessment tool validation studies, job analysis information is used primarily for two purposes: 1) to match a test's meaning and content with the domain of attributes required for job performance, and 2) to develop criterion measures representative of essential components of job performance. The degree of specificity needed for these purposes can vary substantially between psychological tests and physical ability tests.

Physical Domain

Compared to cognitive tests, the structure of physical performance based on actual performance measures and the degree of correlation between physical ability tests is less established. Broadly, there are two types of physical employment tests that function as predictor variables: work samples and basic ability tests (Gebhardt & Baker, 2010).

Work sample tests typically are developed on the basis of content matching between test components and job tasks. To establish the match, detailed information about the types of physical actions taken, their purpose, and their parameters (e.g., heights climbed; distances moved; weights carried; tactics used) is required. Infor-

mation must be collected about the realistic and tactically-appropriate sequences in which these behaviors are performed to build a strong argument that a task-based work sample appropriately samples field performance. Basic ability tests (e.g., push-ups; 1.5-mile run) do not rely on a content match, are based on the notion that they sample job-related physical constructs (e.g., muscular endurance; aerobic capacity), and are associated with evidence demonstrating that their scores are predictive of physical job performance measures. Establishing job relevance requires specification of physical abilities and the degree to which these abilities are anchored to job behavior. Via measurement of these abilities, basic ability tests should ideally activate the specific muscle groups associated with successfully performing job tasks.

As dependent variables, physical criterion measures may take the form of ratings for evaluating generalized physical performance, but may also take the form of job simulations that replicate physical aspects of the job (Gebhardt & Baker, 2010). In the case of either work sample or basic ability tests, if criterion measures are developed for a validation study, job analysis information that specifies discrete physical behaviors and the conditions under which they are performed becomes critical. General work activities are insufficient for this purpose. Adopting a high degree of specificity in job analysis is likely to yield richer job simulations that have high fidelity to the actual work, in turn enhancing the acceptance of the process and providing strong criterion measures to evaluate the physical ability tests.

Psychological Domain
To support use of a cognitive assessment tool, a fairly broad description of job demands may be sufficient: The goal is to establish either the level of information-processing involved in performance, which can be characterized generally for almost any job (Hunter & Hunter, 1984), or that job complexity aligns with an occupational group for which validity evidence is established (Pearlman, Schmidt, & Hunter, 1980). Specific content matching between psychological tests and detailed job requirements often is unnecessary for several reasons. One is that positive correlations between test scores, paired with the general saturation of psychological-based work demands across jobs, reduces the utility of specific job information for identifying relevant predictors (e.g., Murphy, Dzieweczynski, & Zhang, 2009). Second, general frameworks of job performance that reflect correct completion of tasks, dealing with others and managing oneself well, and avoiding counterproductive behavior are considered applicable across occupations, and these domains align with cognitive and personality-based assessment tools (e.g., Borman & Motowidlo, 1993; Campbell & Wiernik, 2015; Cortina & Luchman, 2012).

Third, if developing criterion measures, it is also the case that ratings-based measures of job performance used in applied work are frequently insensitive for differentiating job performance constructs. This is consistent with findings that job performance ratings correlate positively and substantially with cognitive test scores and other assessment methods that tap psychological constructs across many studies of different occupations (e.g., interviews [McDaniel, Whetzel, Schmidt, & Maurer, 1994]; situational judgment tests [McDaniel, Hartman, Whetzel, & Grubb, 2007]). Job simulation criteria are possible for consideration, but appear to be reported infrequently when validating these types of predictors.

Together, these conditions suggest that specificity is a pivotal consideration in physical test validation, and that greater specificity is likely needed in this domain than in the psychological testing domain. If the goal is to support use of psychological test scores, a broad – but systematic – job analysis that establishes general cognitive and personality-oriented job characteristics is usually sufficient. Note that our intent is not to resurrect controversy related to bandwidth-fidelity considerations in predictor—criterion specification; rather, the goal is to raise awareness of nuances at the job analysis stage that will affect subsequent measurement development. Note also that our specification of the psychological domain in this column is not focused on detailed job knowledge or skill that often would be assessed outside of selection for organizational entry.

Physical Task Lists
To illustrate the above discussion, we discuss a simple example and present conceptual guidance that we have found useful in our practice.

In law enforcement, "pursue subjects on foot" encompasses many activities that vary depending upon the terrain, the initial start of the pursuit, the length of the pursuit, and the objective of the pursuit. Without context, the statement does not yield enough information to be useful to accurately describe the nature of the pursuit. Each officer who

is asked to describe the process "pursue" subjectively interprets its meaning from their own point of reference and experience. Differences across officers in the visceral experience of completing the physical activity confound a singular interpretation of the task. In this case, the variety of legitimate interpretations can be problematic for subsequent judgments, such as linkages to abilities or determinations of the level of effort involved in task performance. Without explicitly drawing out the task statements as much as possible, it is highly unlikely officers will assign consistent judgments based on a common interpretation of the task (Fine & Wiley, 1971).

The subtlety lies in determining the appropriate level of detail so as to avoid issues of creating lists so long that their use is prohibitive, such as introducing information processing problems presented by Morgeson and Campion (1997). On the detailed side, tasks can be quantified in terms of mode, frequency, duration, intensity, initial posture, and rest periods (Rayson, 2000). A greater level of detail ensures that abstract physical abilities are anchored to discrete observable job behaviors, ultimately providing guidance on the applicability of an ability or skill (Shephard, 1998).

At the same time, there are limits to avoid going down a specificity-fueled rabbit hole. The concrete and specific actions officers perform as they execute different physical tasks can be described in an infinite number of ways, as there are unique contexts and conditions to which the physical tasks relate. Although there may be an infinite number of ways to describe physical tasks, there are only a limited number of significant patterns of behaviors that describe how officers effectively perform their job. The accomplishment of a specific physical task can be performed under widely varying conditions, occur over a range of difficulty, and involves different specific content. But the fundamental behaviors to achieve successful job performance are parallel, and call for similar kinds and degrees of personal characteristics. The pursuit of a task goal entails adapting behavior to influence a unique environment in order to accomplish the task goal (Miller, Galanter, & Pribram, 1960; Annett & Duncan, 1967). The job analyst must identify these fundamental behaviors and be attuned to their parallelism to create an optimally useful product for a validation study.

Physical Job Analysis in Action

During many of our site visits, we asked officers to make purposeful demonstrations of work activities and training scenarios (e.g., canine handling, technical rope rescue, swift water rescue). The demonstrations helped to identify the physical demands of the work by permitting the observation of the particular movement patterns, muscle group actions, time on task, and effort levels necessary to complete the work activities. We conducted interviews using the Critical Incident Technique (Flanagan, 1954) that involved asking officers to provide examples of demanding physical activities and more routine activities. The identified activities were sufficiently detailed to yield clear intent and consequences to the acts. These techniques were useful for specifying physical tasks at a level of detail sufficient for the purposes introduced above: supporting an appropriate physical construct model and developing content-based work simulation measures.

Each of these techniques will be familiar to a personnel assessment professional. Our take is that the investigation of psychological and physical job domains are complementary, but the nuanced decisions related to specificity make a difference in the interpretability of the resulting information. When investigating a job, the tasks are identified in terms of behaviors the job entails. Simultaneously, the job analyst may view the same tasks from a physicality perspective. This leads to descriptions of physical tasks reflective of situational influences on the physical performance of the job tasks. By viewing a task from different angles, the job analyst gains a richer understanding of the job, and the validation study can capitalize on this understanding.

References

Borman, W. C., & Motowidlo, S. M. (1993). Expanding the criterion domain to include elements of contextual performance. In N. Schmitt, W.C. Borman, and Associates (Eds.), *Personnel selection in organizations* (pp. 71-98). San Francisco: Jossey-Bass.

Campbell, J. P., & Wiernik, B. M. (2015). The modeling and assessment of work performance. *Annual review of organizational psychology and organizational behavior*, 2(1), 47-74.

Cortina, J. M., & Luchman, J. N. (2012). Personnel selection and employee performance. In N. W. Schmitt, S. Highhouse, & I. B. Weiner (Eds.), *Handbook of Psycholog*y, Volume 12: *Industrial and Organizational Psychology* (2nd Ed.; pp. 143-183). Hoboken, Hew Jersey: John Wiley & Sons.

Fine, S. A., & Wiley, W. W. (1971). An introduction to functional job analysis: A scaling of selected tasks from the social welfare field.

Methods for Manpower Analysis No. 4. Kalamazoo, MI: Upjohn Institute for Employment Research.

Flanagan, J. C. (1954). The critical incident technique. *Psychological Bulletin*, 51(4), 327-358.

Gebhardt, D. L., & Baker, T. A. (2010). Physical performance tests. In J. L. Farr & N. T. Tippins (Eds.), *Handbook of employee selection* (pp. 277-298). Routledge.

Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, 96(1), 72-98.

McDaniel, M. A., Hartman, N. S., Whetzel, D. L., & Grubb III, W. L. (2007). Situational judgment tests, response instructions, and validity: A meta-analysis. *Personnel Psychology*, 60(1), 63-91.

McDaniel, M. A., Whetzel, D. L., Schmidt, F. L., & Maurer, S. D. (1994). The validity of employment interviews: A comprehensive review and meta-analysis. Journal of Applied Psychology, 79(4), 599-616.

Morgeson, F. P., & Campion, M. A. (1997). Social and cognitive sources of potential inaccuracy in job analysis. *Journal of Applied Psychology*, 82(5), 627-655.

Murphy, K. R., Dzieweczynski, J. L., & Zhang, Y. (2009). Positive manifold limits the relevance of content-matching strategies for validating selection test batteries. *Journal of Applied Psychology*, 94(4), 1018-1031.

Pearlman, K., Schmidt, F. L., & Hunter, J. E. (1980). Validity generalization results for tests used to predict job proficiency and training success in clerical occupations. *Journal of Applied Psychology*, 65(4), 373-406.

Rayson, M. (2000). Job analysis. In S. Constable and B. Palmer (Eds.), *The process of physical fitness standards development* (pp. 67-100). WPAFB, OH: Human Systems Information Analysis Center.

Shephard, R. (1998). Aging and Exercise. In T.D. Fahey (Ed), *Encyclopedia of Sports Medicine and Science*. Internet Society for Sport Science: http://sportsci.org. 7 March 1998. Retrieved from http://www.sportsci.org/encyc/agingex/agingex.html.

About the ACN

The ACN is the official newsletter of the International Personnel Assessment Council, an association of individuals actively engaged in or contributing to the professional, academic, and practical field of personnel research and assessment. It serves as a source of information about significant activities of the Council, a medium of dialogue and information exchange among members, a method for dissemination of research findings and a forum for the publication of letters and articles of general interest. The Council has approximately 300 members.

The anticipated ACN publication date for the remainder of 2018 is December 14. The respective closing date for submissions is November 16. Submissions for Publication: Prospective authors are invited to send in their articles, research reports, reviews, reactions, discussion papers, conference reports, etc., pertaining to the field of personnel research and assessment. Topics for submission include, but are not limited to:

• Technical
• Practical – lessons learned, best practices
• Legal
• Technology/Tools
• Statistics/Measurement
• Book reviews

Articles and information for inclusion should be submitted directly to the Editor via e-mail at figou@latech.edu. Articles will be accepted only by electronic submission (Word compatible). Submissions should be written according to the Publication Manual of the American Psychological Association, 6th edition. The editor has the prerogative to make minor changes (typographical/grammatical errors, format, etc.); substantial changes will be discussed with the author. Submissions more than 1,500 words should include an abstract of maximum 100 words, preferably with three keywords. If you have questions or need further information, please contact the editor.

The ACN is looking for a new Professional and Scientific Affairs Editor. If you are interested or would like to recommend somebody, please contact Frank Igou at figou@latech.edu or call (318) 278-7154.

IPAC is seeking a new editor for the Assessment Council News (ACN). Interested people should contact Frank Igou at frankigou@gmail.com for information about the position or president@ipacweb.org to be considered.

Editors

Frank Igou
Associate Professor of I-O Psychology
Department of Psychology and Behavioral Sciences
Louisiana Tech University
(318) 278-7154
figou@latech.edu

Jennifer Chapman
Digital Communications Manager
(678) 699-0386
Jennifer9535@gmail.com

Associate Editors

Legal Watch
Ryan O'Leary
Manager
Hiring and Assessment Service
PDRI
3000 Wilson Blvd, Suite 250
Arlington, VA 22201
(202) 321-1204
ryan.oleary@pdri.com

Professional and Scientific Affairs
Dennis Doverspike
Professor of Psychology
Psychology Department
University of Akron
Akron, OH 44325
(330) 972-8372
dennisdoverspike@gmail.com

Advertisement Rates

Advertisement Size	Advertisement Dimensions	Cost per Advertisement	Cost for 4 Issues
Full Page	7.5" X 9.75"	$50	$180
Half Page	7.5" X 4.875	$25	$90
Business Card Size	3.5" X 2"	$12.50	$45