

Trends in Testing: Highlights of a Global Survey

Ann Marie Ryan, Michigan State University

Ilke Inceoglu, Surrey Business School, University of Surrey

Dave Bartram, CEB's SHL Talent Measurement Solutions, & University of Pretoria

Juliya Golubovich, ETS

James Grand, University of Akron

Matthew Reeder, APT Metrics

Eva Derous, Ghent University

Ioannis Nikolaou, Athens University of Economics and Business

Xiang Yao, Peking University

Citation

Ryan, A.M., Inceoglu, I., Bartram, D., Golubovich, J., Grand, J.A., Reeder, M., Derous, E., Nikolaou, I., & Yao, X. (2015). Trends in testing: Highlights of a global survey. In I. Nikolaou and J. Oostrom (Eds.), *Employee recruitment, selection, and assessment: Contemporary issues for theory and practice*. (pp. 136-153). New York, NY: Psychology Press.

This document reflects the manuscript version accepted for publication but may not exactly replicate the final printed version of the article. Please cite the final published version of the paper in subsequent references to this manuscript. The final printed version can be found via its DOI: <https://doi.org/10.4324/9781315742175>

In the mid 1990s, Ryan, McFarland, Baron, and Page (1999) conducted a survey of selection practices globally. Because their study is one of the few published surveys of employer practices, it garnered significant citation in the years that followed. Even though much time has passed since that data collection, a comparable, comprehensive examination of employer practices has not surfaced in the selection area. This chapter provides an overview of a more recent effort to capture trends in testing.

On the surface, hiring practices may have changed dramatically since the mid-1990s due to a number of social, economic, and technological trends. Skill and demographic shifts among labor market occupants and changes in job and occupational requirements have led employers to source applicants for jobs in wider markets (and even globally). Technological developments have facilitated and accelerated staffing processes (Scott & Lezotte, 2012). Greater use of computer- and particularly Internet-based testing has provided organizations with greater efficiency in resource allocation, quicker processing of applicants, and access to a larger pool of potential applicants. Technology has allowed for new and varied ways of presenting assessment content to applicants, but has also heightened concerns regarding test security and potential cheating.

Given that these trends have reshaped hiring and staffing over the past 20 years, this chapter provides an updated description of the practices and policies used by organizations around the world. A 54-item survey on selection practices was translated into 15 languages and data was collected from HR professionals in more than 25 countries. This chapter focuses on trends in test use around the globe; specific country differences are not detailed as sample sizes

varied across countries, with many too small to make specific inferences about trends in individual countries.¹

Survey respondents

A total of 1,197 HR professionals completed an online questionnaire about testing practices and policies. Respondents were sourced via a number of methods targeted specifically at reaching HR professionals. Note that we sought to include HR managers/directors/executives within organizations, not HR consultants or lower level HR employees, and thus our sampling strategy aimed to capture that. Professional associations and in particular selection-related groups were contacted in all the countries selected for inclusion in the study (based on coverage of countries in different clusters in the GLOBE study, House, Hanges, Javidan, Dorman, & Gupta, 2004 as well as practical constraints regarding translation capabilities), and were asked to either email a survey announcement to their mailing lists or to post a notification of the survey on their websites. LinkedIn groups of HR professionals in each targeted country were identified and we posted survey notices in those groups. We also accessed the email list for marketing for a major test publisher, and culled HR manager/director/executive emails from that list for a direct mailing about the survey. Finally, collaborators in several countries had contacts within professional associations and assisted us by distributing the survey link. Thus, it is impossible to calculate a response rate as the true population of HR professionals with internal responsibilities for selection systems is not known.

The largest representation in the sample was from the US (22.9%), Belgium (19.4%), and China (15.4%), with others from Sweden (8.2%), the Netherlands (6.5%), Greece (4.3%),

¹ Analyses of the influence of cultural values on testing practices at a regional level are available from the first author on request.

Portugal (3.4%), France (3.0%) and the United Kingdom (2.0%). Other countries with respondents (less than 2% of total sample) included Italy, Russia, Australia, India, Germany, Hong Kong, Indonesia, Turkey, Brazil, New Zealand, Saudi Arabia, Singapore, Spain, Denmark, and South Africa. Most survey respondents were professionals in the private sector (81%), including professional services (21.2%), manufacturing (17.9%), financial (8.4%), retail (7.1%), health care (6.5%), telecommunications (3.8%), and transportation (3.0%), with smaller numbers in construction, information, utilities, insurance, educational services, hospitality, business consulting, chemical, pharmaceutical, mining, and energy. Most respondents were in an HR function in their organizations but held different types of roles (e.g., HR manager (29.8%), HR executive such as director or vice president (26.3%), HR consultant (8.7%)).

Overview of survey content

Questions addressed several areas:

1. **Decisions to use tests and future plans:** Reasons why organizations elect to use or not use tests, and plans for developing, purchasing, or implementing tests in the future.
2. **Test program description:** How tests are created and used in the hiring process, and characteristics assessed by tests.
3. **Use of technology:** Use of adaptive testing, use of supervision and other security measures when testing applicants, reasons for choosing to administer tests without supervision, differences in supervision practices by test type, estimates of cheating and of disqualification of applicants for cheating, and security and data protection practices.

4. **Test policies and practices:** Frequency and type of feedback provided to applicants, reasons for not providing feedback, retesting policies, global testing practices such as use of standardized testing practices across countries and practices associated with administering tests in multiple languages (e.g., translation, psychometric adequacy, evaluation), and metrics used to monitor the effectiveness of tests (e.g., job performance, attrition, hiring process efficiency, return on investment).

Note that we focused specifically on testing rather than other aspects of a hiring process (e.g., interviewing, recruiting, applicant tracking) in the interest of keeping the survey at a reasonable length while gathering sufficient detail on specific current trends. We defined test for respondents as “any standardized assessment instrument other than an interview or resume review that is designed to evaluate whether a job applicant possesses certain qualities and characteristics (e.g., knowledge, skill, traits).”

In the following sections we detail key findings in each of these areas.

Decisions to use tests and future plans

Researchers have long been interested in understanding why employers decide to use or not use different selection tools (Terpstra & Rozell, 1997; Wilk & Capelli, 2003). About 60% of respondents said their organizations typically use tests for selecting entry-level management employees. Of particular interest is why organizations choose not to use tests (see Table 1). Consistent with earlier research on the predictors of selection tool use (König, Klehe, Berchtold & Kleinmann, 2010), cost and the extent to which use of tests is common practice for targeted jobs or locations were of relatively greater concern than legal considerations. But, in contrast to

the earlier research (König et al., 2010), which had found perceived tool validity to have modest importance, many of the reasons for not using standardized assessments indicated by our respondents seem to represent a lack of belief in or knowledge about the value of tests (e.g., preferences for other methods, inability to obtain buy-in, unable to assess return on investment (ROI)). Thus, continued concerted efforts by testing professionals to educate and inform HR managers about the value of tests seem warranted. Klehe (2004) provides a framework that outlines the many institutional pressures (internal markets, industry norms) that affect organizations' willingness to adopt selection procedures; analyzing these factors might enable testing professionals to garner a better understanding of when and why organizations may not respond to efforts to educate decision makers on the value of testing in particular contexts.

The literature also suggests that some types of tests may not be adopted because of tool-specific concerns, such as faking on personality tests (Rothstein & Goffin, 2006) and resources needed for building and administering simulations (Whetzel, McDaniel, & Pollack, 2012) so we also asked about reasons for not using specific assessment types (e.g., cognitive ability, personality, simulations). In most cases, top reasons were beliefs that the particular skill/ability assessed was not needed for the job or that the test would overlap with other parts of the hiring process (e.g., interview). While it is true that an interview can be used to assess many things (Huffcutt, Conway, Roth, & Stone, 2001), traditional (unstructured) interviews have low validity (e.g., Conway, Jako, & Goodman, 1995). Further, even when more valid, structured interviews are used, depending on the constructs they are designed to assess, additional assessments (e.g., personality or cognitive test) may provide incremental validity (Berry, Sackett, & Landers, 2007). Providing practitioners with a clearer understanding of the intercorrelations of various testing tools and interviews in understandable language might enable individuals to better

understand the degree of overlap; translating concepts such as incremental validity into language familiar to organizational stakeholders would also be valuable to increasing test adoption and selection system effectiveness (Boudreau, 2012).

Finally, among respondents whose companies do not currently use tests, approximately 40% indicated that they do plan on developing, purchasing, or implementing tests for hiring in the next three years. Based on this data, one might forecast an increased use of testing tools by organizations, as would fit with the trends noted earlier regarding technology and the ease of test use.

We also asked those who already used testing in some capacity why they had adopted tests in their hiring processes (see Table 2). Validity/effectiveness, fairness, and perceived value are the top three factors that influence companies' decisions to use tests. This again highlights how important persuading HR decision-makers of the value of testing is to adoption. These findings are interesting as König et al. (2010) found that validity was only a modest consideration in adoption decisions; our broader sample and use of the term "effectiveness" which may encompass different types of evidence than formal evaluation may explain the differences. Legal/political considerations, reducing time required of applicants, and reinforcing the employer brand were the top three reasons "not important" for decisions to test. König et al. (2010) had likewise found legal and organizational self-promotion to be modest predictors of test adoption. The rest of this chapter focuses on this subsample (N = 766) of test users and details how they use testing.

Test program descriptions

Tests in use were more commonly created by individuals external to the organization (50.8%) or through collaboration with external individuals (41.8%) than solely by those working within the organization (19.8%).² Companies used tests at different stages of the selection process (beginning (20.9%), intermediate (50.7%), end (23.3%) total N =756). Tests were typically used along with other tools to make selection decisions as only 2% of respondents reported using tests as the only tool in selection. Personality, abilities, and leadership competencies were the most common characteristics assessed by tests. Interests were among the least commonly assessed (see Table 3).

Of particular interest is that although most companies use test scores in a relatively formal manner, either by combining test scores and interview ratings in a standardized manner to make decisions (43.1%; N = 745) or by using tests as screeners before interviews (25.1%), a substantial portion of respondents (27.1%) indicated that test scores and interpretive information are provided to hiring managers, who make decisions. It is important to consider how much bearing objective test scores have on managerial decisions when scores are used in this less formal way, particularly when a manager's subjective intuition about a candidate is at odds with the individual's scores. Managers' implicit beliefs can inhibit their willingness to use test information in hiring (Highhouse, 2008) and some managers have explicit preferences for intuition-based hiring (Lodato, Highhouse, & Brooks, 2011). Providing hiring managers with some degree of control (e.g., you cannot hire a candidate with scores below a certain level and you are cautioned about hiring others in a "yellow" zone, but are free to choose those with a "green" test score) may lessen their resistance to additional structure in the hiring process.

² Note that some respondents indicated more than one response for this question (e.g., some of the organization's tests were created externally while others were created collaboratively). Total N = 754.

Considering ways to allow managers to feel their preference is met while simultaneously structuring elements of the process to ensure test information is appropriately weighed is an area in need of further research (Lievens & De Paepe, 2004).

Use of Technology

Because the greatest changes in selection practice since the 1990s appear to be technology-linked, we focused much of our survey on questions on the use of technology in testing.

Assessment content. While the general trend is toward increased use of technology in the hiring process, there are differences across employers in the adoption of technology. Some methods of assessment were more likely to be computerized than paper and pencil (assessments of cognitive ability, language capability, personality (work styles), interests, integrity, and situational judgment) and others more likely to not be computerized (job knowledge test, simulation test (in-basket, role play)). Across the test types we asked about, an average of 14% of respondents indicated that their organizations test in *both* paper and pencil and computerized formats. Approximately 87% of respondents have considered or are currently considering using computerized tests in their organizations (N = 542).

Technology has been widely advocated as a means of expanding what is assessed and how it is assessed (e.g., new KSAs, new formats). As Table 4 indicates, drag-and-drop items and video/multimedia are more commonly used elements in computerized testing than animation, interactive voice response, and avatars despite how much the latter are touted as benefits of computerized assessments (Reynolds & Dickter, 2010; Scott & Lezotte, 2012).

Proctoring practices. Computerized tests can be administered in either a supervised or an unsupervised setting. Related to the latter, a major concern among organizational psychologists

has been the use of unproctored tests (Tippins, 2009; Tippins et al., 2006). Among those whose companies use computerized testing, 40.2% (N = 691) indicated using unproctored testing for all their computerized testing, or using unsupervised testing depending on hiring process stage (23.6%) or geographic location of applicant/job (16.1%). A minority of respondents (20.1%) said that all computerized testing was supervised. As Tippins (2009) noted, “the UIT [unproctored internet testing] train has left the station” (p. 4) and debate about the viability or ethicality of the practice needs to be replaced by research on how to improve practices. The primary reason driving decisions to administer computerized tests in unsupervised settings appears to be the desire to make the process convenient for applicants (65.9%; N = 531). Other frequently stated reasons are cost effectiveness (54.4%) and convenience for hiring managers (53.9%), easier assessment of a larger applicant pool (51.8%), and reduction in time-to-hire (51.8%). As this list shows, unproctored testing is adopted for efficiency reasons (see Scott & Lezotte, 2012); it is therefore incumbent upon psychologists to ensure that greater efficiency does not necessarily mean lower quality/effectiveness. This has been a particular concern of testing standards groups (see Naglieri et al., 2004; International Testing Commission, 2006).

In response to calls for a better understanding of proctoring practices in employment testing (Arthur & Glaze, 2011; Drasgow, Nye, Guo, & Tay., 2009;), we asked a number of more specific questions about how companies use unsupervised tests. As shown in Table 5, companies’ practices of unsupervised testing vary somewhat by type of test. Note that we asked about supervision for both paper and pencil and computerized tests. As Drasgow et al.(2009) have noted, it is wrong to automatically assume that proctoring occurs when testing is via paper and pencil, and our data support that. However, unsupervised testing is more likely when the tests are computerized.

Personality and background data assessments are most frequently administered unsupervised. Tests that evaluate candidates' cognitive ability, knowledge, and judgment are somewhat less frequently administered unsupervised but still used this way fairly often. Implications of administering cognitive tests, job knowledge tests, and other assessments on which a candidate could cheat (e.g., have a substitute take the test, use an advisor, share the test with others) unsupervised has been of greatest concern to researchers (Arthur & Glaze, 2011). These concerns were expressed by respondents in our study as well. Of those using *supervised* computerized testing (N = 378), 51.6% reported that the risk of cheating was too great and 36.5% reported concern over test content dissemination. Only 33.3% supervised because of concern over applicant comfort with technology and only 29.1% reported the Internet was not universally available for their applicant pool.

Table 6 shows that strategies for delivery of items are evolving. About 35 of respondents (N = 666) said their organizations use either adaptive computerized tests or randomly selected items from a larger pool (44% said they did not use adaptive tests and 20% said they did not know). Varying test content (e.g., using different items or different forms) is a relatively uncommon practice. About 54% of respondents indicated their companies use fixed tests. The promise of computerized testing is still to be delivered, as many organizations apparently simply have created page-turner versions of paper and pencil test items (Potosky & Bobko, 2004).

Security measures and data protection. We asked about companies' security measures to better understand the extent to which companies employ test administration practices that may minimize the chances of cheating and tests becoming compromised. The most frequently used security measure with unproctored computerized testing is adhering to time limits (see Table 7). Research indicates that administering speeded tests can help to minimize cheating as time

constraints limit opportunities for these behaviors (Arthur, Glaze, Villado, & Taylor, 2010). A substantial number of respondents (40%) also indicated that their companies use warnings, which have likewise been recommended in the research literature for minimizing intentional distortion (e.g., Hough, 1998). Although testing guidelines advocate the use of verification testing (i.e., administration of a proctored confirmation test to those initially assessed remotely to detect cheating; Naglieri et al., 2004; ITC, 2006), fewer than 20% of respondents indicated using verification testing; however, many companies may be using tests such as personality measures, where verification makes less sense. Finally, consistent with Arthur et al.'s (2010) recent observation, few companies seem to be using technological innovations for monitoring candidates (e.g., webcams, keystroke analyses). Of the technological tools we asked about, preventing backtracking and other computer applications from running were the most commonly used.

In addition to asking about security measures specific for unproctored computerized testing, we asked about security measures for paper and pencil testing as well as for supervised computerized testing. As Tables 8 and 9 indicate, the most frequently used security measure for both paper and pencil testing and supervised computerized testing is following test procedures and adhering to time limits for tests. Companies using paper and pencil tests seem least concerned with test materials going missing (only 31% count and keep track of test materials). In summary, Drasgow et al.'s (2009) assertion that test security is not necessarily strong for paper and pencil testing is supported by the survey results.

Approximately half of the respondents believe that somewhere between 1 and 20% of applicants cheat or misrepresent themselves on their organizations' tests, regardless of the test format (paper and pencil, supervised computerized, unsupervised computerized; see Table 10).

As Arthur and Glaze (2011) noted, the expectation is not that there is no cheating in proctored settings but that rates may increase in unproctored settings. Unsupervised computerized tests were associated with the highest uncertainty among respondents about the amount of cheating that happens, and this format is seen as presenting most risk (only 9.3% of respondents thought applicants could not cheat).

Among those choosing to use unproctored computer tests, the risk that applicants may cheat and the uncertainty regarding the extent to which they actually do apparently does not outweigh the efficiency gained by administering computerized tests in an unsupervised setting (recall that efficiency considerations were the top drivers of the decision to adopt this method of testing). The majority of respondents were willing to tolerate up to 20% of applicants cheating on an unsupervised computerized test (71.9% would not stop using the test for selection). Cheating on UITs may not be quite that high, however. For example, Arthur et al. (2009) estimate 7.7% of their sample cheated on a cognitive ability unproctored computerized test.

Interestingly, respondents actually indicated similar attitudes about cheating regardless of method of test administration. The majority of respondents were also willing to tolerate up to 20% of applicants cheating on a paper and pencil test (67.4% would not stop using) or a supervised computerized test (76.8% would not stop using) (compare to percentages cited above for unproctored computerized tests). However, Arthur and Glaze (2011) note, the real concern is not with the number of cheaters but with their distributional placement and relation to cut scores. That is, it matters less what total percent cheat and more what percentage of cheaters receive a passing score when they would not have otherwise or who end up ranking higher than honest test takers who they would not otherwise have surpassed. Further, the majority of respondents said that their organizations either never or very rarely had to disqualify applicants

for cheating on paper and pencil tests (85%), supervised computerized tests (90%), or unsupervised computerized tests (93%).

The most common data protection strategy was allowing only relevant staff to access test data (93.1%; N = 664-729). Other strategies were to use firewalls and password protections (81.7%), to have physical security where data is stored (66.2%), to have regular data backups (65.8%), to ensure data is protected in electronic transit (e.g., by encryptions; 55.0%), and to have disaster recovery plans in place (37.8%). Respondents also seemed to be relatively less familiar with certain data protection strategies. For example, about 43% of respondents were not aware of whether their organizations have disaster recovery plans. This overall lack of attention to data protection is disconcerting, particularly given the European Union Privacy Directive (1998) and the US Safe Harbor Provisions (2000), which attempt to set guidelines for the protection of personal data and test data (Reynolds & Dickter, 2010).

Test policies and practices

Researchers have been interested in the effects of test feedback (or lack thereof) on applicant perceptions (see London & McFarland, 2010 for summary of research). About half of the respondents (51.3%; N = 745) said their companies almost always or always provide applicants with feedback on test results. A minority (8.7%) of respondents indicated that their companies never provide feedback to applicants on test results. About 65% of respondents (N = 676) indicated that their companies explain to applicants how to interpret a test score and 50.7% at least provide applicants with their test scores, while 45.9% provide pass/fail feedback. Letting applicants know how they did relative to others is relatively uncommon (23.7%) as is providing other normative score information (33.1%). The most common reason for not providing

feedback (N = 59) to applicants is time constraints (20.3%), followed by lack of benefit to the organization (18.6%) and concerns about legal liabilities (18.6%). Cost is typically not a factor (5.1%). Applicant complaints about not getting timely feedback are common (Gilliland, 1995) so one question is how quickly applicants receive this feedback. Another factor behind applicant concerns is that they may be desiring more specific feedback than is typically provided. Also, complaints about lack of feedback may be related more to interviewing than testing processes.

Retake policies have also been a focus of considerable research (Hausknecht, Halpert, DiPaolo, Moriarty, & Gerrard, 2007; Lievens, Buyse, & Sackett, 2005; Schleicher, van Iddekinge, Morgeson, & Campion, 2010), with studies indicating score increase upon retesting, which may affect validity in some cases. Surprisingly, 77% of respondents (N = 739) said their companies do not allow job applicants to retake assessments if they initially failed or were not hired. It may be that respondents interpreted our question to be about immediate retests rather than retesting after a set interval, which is part of many testing policies. Applicants are most often allowed to retake cognitive ability tests (58.2%; N = 92). Retesting is rare for integrity tests and interest assessments (less than 15% allow; N = 22). When companies allow retesting, applicants are more likely to take the same exact test than a different version of the test for assessments of background data, interests, personality, and situational judgment. Applicants are more likely to take a different version of the test than the same exact test for assessments of cognitive ability, integrity, and language capability. Applicants are about equally likely to take the same exact test and a different version of the test for job knowledge and simulation tests. While one can appreciate cost concerns of alternate forms (see for example Lievens & Sackett, 2007 on SJT alternate form development), fielding only one version has test security risks. Note,

though, that few respondents actually answered the question about retesting opportunities (Ns from 17-81).

The majority (66%) of respondents (N = 759) said their organizations do not administer tests globally. The majority (71%) of respondents whose organizations do administer tests globally (N = 250) said they test in more than one language. It is typical for multinational testing practices (N = 247) to either be the same across countries (39.7%) or to include a combination of custom and standardized processes (39.3%), rather than use different processes across countries (21.1%). Most companies (62.2%) that administer tests in multiple languages let their assessment vendors handle matters pertaining to translation; indeed, in forecasting the future of selection Ryan and Ployhart (in press) noted that the trend for outsourcing of selection tool development and research is likely to continue to grow. Table 11 details other practices when testing across languages; it is clear that not all recommended practices are being followed (International Test Commission, 2006).

Finally, we asked respondents about how they evaluated the effectiveness of testing programs. As Table 12 indicates, the most frequently monitored metric is job performance of those hired (70.6%). ROI for tests is calculated relatively infrequently (19.2%). Note that “monitoring” a metric does not necessarily mean that organizations are engaged in ongoing, rigorous validation studies.

Research Implications

Throughout this chapter we have noted some of the practical implications of our findings, but there are also important implications for selection researchers. First, in relation to the practical implications, is the need for research on methods of “messaging” assessment practices so that

they are more likely to be adopted. As noted earlier, researchers like Boudreau (2012) have suggested we retool in terms of how we try to communicate concepts like utility; further research in this area can only be helpful. The challenge for selection experts is how to improve the communication of this information so that it is readily available on a global basis to interested HR practitioners and managers and that it is in a format that is easily understood and used.

Second, research on the adoption of innovation and technology has been applied to some extent to selection contexts (Konig et al., 2010), but there is ample room for further theoretical development in this regard. Third, the security methods examined here were chosen based on what is considered good practice, but research as to the usefulness and effectiveness of their employment in deterring cheating is still limited.

Finally, in this paper we report findings on assessment use from a global sample. Research to tease apart what factors affect adoption, such as legal environment, unionization, country economic conditions, educational systems, etc... has not been particularly systematic, largely due to the challenges in obtaining adequate samples from different countries. It may behoove researchers to form consortia and to focus efforts on specific factors that vary globally as considerations in sampling so as to better design studies that can adequately test the role of these factors in selection practice adoption.

Limitations

As with any effort, this survey was not without limitations. As noted earlier, we were challenged to identify appropriate respondents (HR managers and executives with responsibilities for selection programs), particularly in certain countries. This led us to use professional associations and web groups as a primary means of sourcing respondents, and made

response rates incalculable (i.e., we did not have access to total numbers of members, or total number of views of web pages). Further, those who do not test may be less likely to respond to such a survey. We were not able to access multiple respondents per organization to provide us with reliability information, although most questions were designed to be objective. Collecting data globally also presents challenges in that we did not have access to associations or contacts in certain locations, and we only possessed resources to have professional translations in 15 languages.

Summary Recommendations

Based on the findings of this survey, we see a number of directions for organizational psychology research and practice:

- 1) Reasons for using or not using testing were tied to the value of testing, suggesting that continued work to document and especially to communicate the value of testing should be a focus of research and practice efforts. In particular, enhanced communication regarding the incremental validity of testing may be important to adoption decisions.
- 2) Companies have taken advantage of the availability of technology to move away from using a paper and pencil format for most types of tests. However, most do not seem to be using the capabilities provided by recent technological advancements to the extent possible, in that less than half of respondents indicated using various elements made feasible by computerized tests (e.g., video/multimedia, avatars, adaptive testing). Researchers and practitioners can focus efforts on enhancing these technological advances and promoting greater use.

- 3) Test security practices do not seem to be widely or fully employed for paper and pencil or supervised computerized testing let alone for unproctored computerized tests. The value of a selection system can be completely degraded by poor security so attention by practitioners to communicating the importance of security and data protection, as well as attention to means of making security measures easy to implement may help. Development of alternate forms in cases where adaptive pools are not in use should also be a focus given retesting policies. Note that this lack of attention to security may be due to beliefs that not many individuals cheat, willingness to tolerate a certain rate of cheating, and the rarity of detecting cheaters.
- 4) Global testing programs are likely to increase given the globalization of business, suggesting a need for greater attention to international testing standards. Many of the advocated practices for using testing worldwide did not appear to be followed.
- 5) Organizations increasingly track metrics that may be used to evaluate selection systems; further work to establish high quality evaluation programs may even further support the value of test use in selection.

References

- Arthur, W., & Glaze, R. M. (2011). Cheating and response distortion on remotely delivered assessments. In N.T. Tippins & S. Adler (Eds.). *Technology-enhanced assessment of talent* (pp. 99-152) San Francisco: Jossey-Bass.
- Arthur, W., Glaze, R. M., Villado, A.J., & Taylor, J. E. (2009). Unproctored internet-based tests of cognitive ability and personality: Magnitude of cheating and response distortion. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, 2, 39-45.
- Arthur, W., Glaze, R.M., Villado, A.J., & Taylor, J.E. (2010). The magnitude and extent of cheating and response distortion effects on unproctored internet-based tests of cognitive ability and personality. *International Journal of Selection and Assessment*, 18, 1-16.
- Bauer, T. N., Maertz, C. P., Dolen, M. R., & Campion, M. A. (1998). Longitudinal assessment of applicant reactions to employment testing and test outcome feedback. *Journal of Applied Psychology*, 83, 892-903.
- Berry, C. M., Sackett, P. R., & Landers, R. N. (2007). Revisiting interview-cognitive ability relationships: Attending to specific range restriction mechanisms in meta-analysis. *Personnel Psychology*, 60, 837-874.
- Boudreau, J. W. (2012). "Retooling" Evidence-Based Staffing: Extending the Validation Paradigm Using Management Mental Models. In N. Schmitt (Ed.), *The Oxford handbook of personnel assessment and selection* (pp. 793-813). New York: Oxford University Press, Inc.

- Conway, J. M., Jako, R. A., & Goodman, D. F. (1995). A meta-analysis of interrater and internal consistency reliability of selection interviews. *Journal of Applied Psychology, 80*(5), 565-579
- Drasgow, F., Nye, C. D., Guo, J., & Tay, L. (2009). Cheating on proctored tests: The other side of the unproctored debate. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 2*(1), 46-48.
- European Commission. (1995, Nov. 23). European Privacy Directive (EU Directive 95/46/EC). *Official Journal of the European Communities*, No L. 28.
- Gilliland, S. W. (1995). Fairness from the applicant's perspective: Reactions to employee selection procedures. *International Journal of Selection and Assessment, 3*(1), 11-19
- Hausknecht, J. P., Halpert, J. A., Di Paolo, N.T., Moriarty, N.T., & Gerrard, M. O. (2007). Retesting in selection: A meta-analysis of practice effects for tests of cognitive ability. *Journal of Applied Psychology, 92*(2), 373-385
- Highhouse, S. (2008). Stubborn reliance on intuition and subjectivity in employee selection. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 1*(3), 333-342.
- Hough, L. M. (1998). Effects of intentional distortion in personality measurement and evaluation of suggested palliative. *Human Performance, 11*, 209-244.
- House, R. J., Hanges, P. J., Javidan, M., Dorfman, P.W., & Gupta, V. (2004). *Culture, leadership, and organizations: The GLOBE study of 62 societies*. Thousand Oaks, CA: Sage.

- Huffcutt, A. I., Conway, J. M., Roth, P. L., & Stone, N. J. (2001). Identification and meta-analytic assessment of psychological constructs measured in employment interviews. *Journal of Applied Psychology, 86*, 897-913.
- International Test Commission (2006). International guidelines on computer-based and internet-delivered testing. *International Journal of Testing, 6* (2), 143-171
- Klehe, U.-C. (2004). Choosing how to choose: Institutional pressures affecting the adoption of personnel selection procedures. *International Journal of Selection and Assessment, 12*, 327-342.
- König, C. J., Klehe, U.-C., Berchtold M., & Kleinmann, M. (2010). Reasons for being selective when choosing personnel selection procedures. *International Journal of Selection and Assessment, 18*, 17-26.
- Lievens, F., Buyse, T., & Sackett, P. R. (2005) Retest effects in operational selection settings: Development and test of a framework. *Personnel Psychology, 58*, 981–1007.
- Lievens, F. & De Paepe, A. (2004). An empirical investigation of interviewer-related factors that discourage the use of high structure interviews. *Journal of Organizational Behavior, 25*, 29-46.
- Lievens, F. & Sackett, P. R. (2007). Situational judgment tests in high-stakes settings: Issues and strategies with generating alternative forms. *Journal of Applied Psychology, 92*, 1043-1055.
- Lodato, M. A., Highhouse, S., & Brooks, M. E. (2011). Predicting professional preferences for intuition-based hiring. *Journal of Managerial Psychology, 26*, 352-365.

- London, M., & McFarland L.A. (2010). Assessment feedback. In J.L. Farr & N.T. Tippins (Eds.). *Handbook of employee selection* (pp. 417-438). New York: Routledge
- Naglieri, J. A., Drasgow, F., Schmit, M., Handler, L., Prifitera, A., Margolis, A., et al. (2004). Psychological testing on the internet: New problems, old issues *American Psychologist*, 59(3), 150-162.
- Potosky, D., & Bobko, P. (2004). Selection testing via the Internet: Practical considerations and exploratory empirical findings. *Personnel Psychology*, 57, 1003-1034.
- Reynolds, D. H., & Dickter, D. N. (2010). Technology and employee selection. In J.L. Farr & N.T. Tippins (Eds.), *Handbook of employee selection* (pp. 171-194). New York: Routledge.
- Rothstein, M. G., & Goffin, R. D. (2006). The use of personality measures in personnel selection: What does current research support? *Human Resource Management Review*, 16, 155-180.
- Rothstein, M. G., & Goffin, R. D. (2006). The use of personality measures in personnel selection: What does current research support? *Human Resource Management Review*, 16(2), 155-180.
- Ryan, A. M., McFarland, L., Baron, H., & Page, R. (1999). An international look at selection practices: Nation and culture as explanations for variability in practice. *Personnel Psychology*, 52(2), 359-391.
- Ryan, A. M. & Ployhart, R. E. (in press). A century of selection. *Annual Review of Psychology*.

- Schleicher, D. J., van Iddekinge, C. H., Morgeson, F. P., & Campion, M. A. (2010). If at first you don't succeed, try, try again: Understanding race, age, and gender differences in retesting score improvement. *Journal of Applied Psychology, 95*, 603-617.
- Scott, J. C., & Lezotte, D.V. (2012). Web-based assessments. In N. Schmitt (Ed.), *The Oxford Handbook of Personnel Assessment and Selection* (pp. 485-516). Oxford University Press.
- Terpstra, D., & Rozell, E. (1997). Why some potentially effective staffing practices are seldom used. *Public Personnel Management, 26*, 483-495.
- Tippins, N. T. (2009). Internet alternatives to traditional proctored testing: Where are we now? *Industrial and Organizational Psychology: Perspectives on Science and Practice, 2*, 2-10.
- Tippins, N.T., Beaty, J., Drasgow, F., Gibson, W. M., Pearlman, K., Segall, D. O., & Shepherd, W. (2006). Unproctored internet testing in employment settings. *Personnel Psychology, 59*, 189-225.
- Van Iddekinge, C. H., Roth, P. L., Putka, D. J., & Lanivich, S. E. (2011). Are you interested? A meta-analysis of relations between vocational interests and employee performance and turnover. *Journal of Applied Psychology, 96*(6), 1167-1194.
- Whetzel, D. L., McDaniel, M. A., & Pollack, J. M. (2012). Work Simulations. In M. A. Wilson, W. Bennett, S. G. Gibson, & G. M. Alliger (Eds.), *The handbook of work analysis: Methods, systems, applications and science of work measurement in organizations* (pp. 401-418). New York: Routledge.

Wilk, S. L., & Cappelli, P. (2003). Understanding the determinants of employer use of selection methods. *Personnel Psychology*, 56, 103-124.

Table 1: Reasons for not using tests

	<u>Percentage of respondents</u>
Prefer own methods of testing (e.g., interviewing, résumé or CV sifting)	60.4%
Too expensive	37.7%
Too uncommon a practice for this type of job	30.3%
Inability to obtain internal buy-in or support to use testing	30.3%
Unable to effectively implement (e.g., lack technology or personnel to administer)	28.4%
Unaware of tests that would assess what we are looking for	27.0%
Not enough candidates to justify cost	26.8%
Adds too much to total time-to-hire	25.1%
Unable to calculate ROI of using tests	23.5%
Too uncommon a practice in locations where we hire	19.9%
Overlaps too much with other methods of assessing candidates	15.0%
Prior negative experiences with testing	13.4%
Applicants can cheat or fake answers too easily	11.7%
Poses too great a legal risk to use	10.4%
Insufficient support/training from vendor/provider	5.2%
Other	18.3%

Respondents = 366

Note: Respondents could select more than one answer

Table 2: Factors influencing decisions to test

	<u>Very important</u>
Validity/effectiveness	82.9%
Fairness	67.9%
Perceived value	61.7%
Ease of use by organization	55.5%
Prior positive experience	54.5%
Ease of use by applicants	34.5%
Ability to reduce applicant pool	32.6%
To reduce time required of hiring managers	31.4%
To reduce time to hire	30.4%
Reinforces employer brand	24.8%
Legal/political considerations	21.9%
To reduce time required of applicants	17.9%

Respondents = 725-738

Note: Respondents rated different reasons on importance

Table 3: Characteristics assessed by tests

	<u>Percentage</u> <u>of</u> <u>respondents</u>
Personality (e.g., conscientiousness, adaptability, work styles)	84.5%
Abilities (e.g., math, verbal, language)	81.6%
Leadership competencies	65.3%
Social skills (e.g., interpersonal skill, social perceptiveness)	59.6%
Motivation (e.g., achievement orientation)	57.7%
Administrative skills (e.g., planning, organizing)	53.8%
Knowledge (e.g., job specific technical knowledge)	51.8%
Work values (e.g., autonomy)	48.9%
Experience (e.g., background)	22.5%
Interests	18.9%
Other	5.6%

Respondents = 755

Note: Respondents could select more than one answer

Table 4: Elements used in computerized testing

-	<u>Already use</u>	<u>Intend to use</u>
Drag-and-drop items	46.3%	32.2%
Video/multimedia in test item content	44.2%	51.0%
Video/multimedia images in test instructions	41.7%	49.8%
Audio	30.4%	31.4%
Animation in test content	26.9%	32.9%
Interactive voice response	9.9%	25.5%
Avatars (computer generated visual representation of the candidate)	8.1%	21.6%

Respondents = 283

Note: Respondents could select more than one answer

Table 5: Use of unsupervised tests

	<u>Unsupervised</u> <u>paper & pencil</u>	<u>Unsupervised</u> <u>computerized</u>
Personality assessment (work styles)	40.2%	79.7%
Background data	49.4%	61.2%
Cognitive ability test	20.3%	59.8%
Interests assessment	36.7%	56.1%
Integrity test	33.3%	54.2%
Language capability test	26.4%	50.9%
Situational judgment test	23.7%	46.8%
Job knowledge test	25.8%	45.9%
Simulation test (in-basket, role play)	25.2%	40.5%

Respondents = 39 – 531

Note: Respondents could select more than one answer

Table 6: Strategies used when administering unsupervised computerized tests

	<u>Percentage of respondents</u>
-	
Use a fixed test that does not change	53.9%
Randomize order of items for each test administration	30.2%
Restrict when participants can take the computerized test (e.g. , a specific date, time, place, etc)	23.7%
Periodically refresh item content (i.e. , replace items with similarly calibrated ones from an item bank, replace the entire test with an alternate version)	20.4%
Create a unique version of the test for each applicant based on responses to each item (computer adaptive testing)	17.6%
Create a unique version of the test for each applicant using randomly selected items from a large item bank	17.0%
Create a new version of the test for a job opening using randomly selected items from a large item bank	14.6%
Rotate among several different forms of the test across applicants	11.7%

Respondents = 460

Note: Respondents could select more than one answer

Table 7: Security measures used when administering unsupervised computerized tests

	<u>Percentage of respondents</u>
Strict time limits	59.3%
Use of warnings regarding cheating	40.0%
No backtracking	32.1%
Disabling other applications on the computer	19.5%
Use of supervised confirmation or verification testing	18.3%
Use of honesty certificates that require examinees to certify they will not cheat	13.6%
Use of webcams	6.7%
Use of keystroke analyses	4.7%
Other	7.7%

Respondents = 405

Note: Respondents could select more than one answer

Table 8: Security measures used when administering paper & pencil tests

	<u>Percentage</u>
	<u>of</u>
	<u>respondents</u>
Strictly follow test procedures and adhere to administration time limits	71.8%
Prohibit copying or reproducing test materials	59.8%
Allow access to tests only to personnel with a legitimate need	54.9%
Always use properly trained test administrators and proctors	53.1%
Never leave applicants unsupervised with access to secure test materials	49.6%
Provide testing accommodations only to those eligible to receive them	46.3%
Store test materials in a secure, locked area	44.1%
Count and keep track of the number of secure test materials	31.0%

Respondents = 510

Note: Respondents could select more than one answer

Table 9: Security measures used when administering supervised computerized tests

	<u>Percentage of respondents</u>
Strictly follow test procedures and adhere to administration time limits	65.9%
Password protect test materials	57.7%
Always use properly trained test administrators and proctors	56.2%
Allow access to tests only to personnel with a legitimate need	55.1%
Prohibit copying or reproducing test materials	54.6%
Never leave applicants unsupervised with access to secure test materials	46.5%
None of the above/Other	3.1%

Respondents = 381

Note: Respondents could select more than one answer

Table 10: Beliefs about the percentage of applicants who engage in some form of cheating or misrepresentation on the organization’s tests

	<u>Paper & pencil tests</u>	<u>Supervised computerized tests</u>	<u>Unsupervised computerized tests</u>
-			
Do not know	24.6%	29.0%	35.1%
0% (Not possible for applicants to cheat)	16.4%	19.3%	9.3%
1-5%	25.5%	28.0%	31.4%
6-10%	14.9%	13.5%	12.3%
11-20%	10.2%	7.1%	7.3%
21-30%	6.9%	2.5%	3.7%
More than 30%	1.5%	0.5%	0.7%

Respondents = 393 – 549

Table 11: Practices used when dealing with tests that are administered in multiple languages

	<u>Percentage of respondents</u>
Our assessment vendor handles all matters pertaining to translation	62.2%
Review by end users in countries of use	35.4%
Back translation procedures to ensure accuracy	29.3%
Development of separate norms for different country/language groups	29.3%
Revision beyond translation to accommodate cultural differences/nuances	24.4%
Psychometric assessments of measurement equivalence	21.3%
Separate validation studies for each translation	14.0%
Other	4.3%

Respondents = 164

Note: Respondents could select more than one answer

Table 12: Regularly monitored metrics

	<u>Percentage of respondents</u>
Job performance of those hired	70.6%
Attrition rates of those hired	45.3%
Opinions of key internal stakeholders on effectiveness of selection tools	41.7%
Process efficiency (e.g. , cost pre-hire, time to hire)	39.0%
Views of applicants on our selection process	37.4%
Pass/fail rates	35.5%
Return on investment for testing (ROI)	19.2%
Other	2.7%

Respondents = 677

Note: Respondents could select more than one answer