Using Technology in Personnel Assessment

Nancy T. Tippins
CEB Valtera
July 23, 2013

ROAD MAP FOR THE PRESENTATION

What is technology-enhanced assessment ?

What's good and bad about using technology in personnel assessment ?

What questions should be asking?

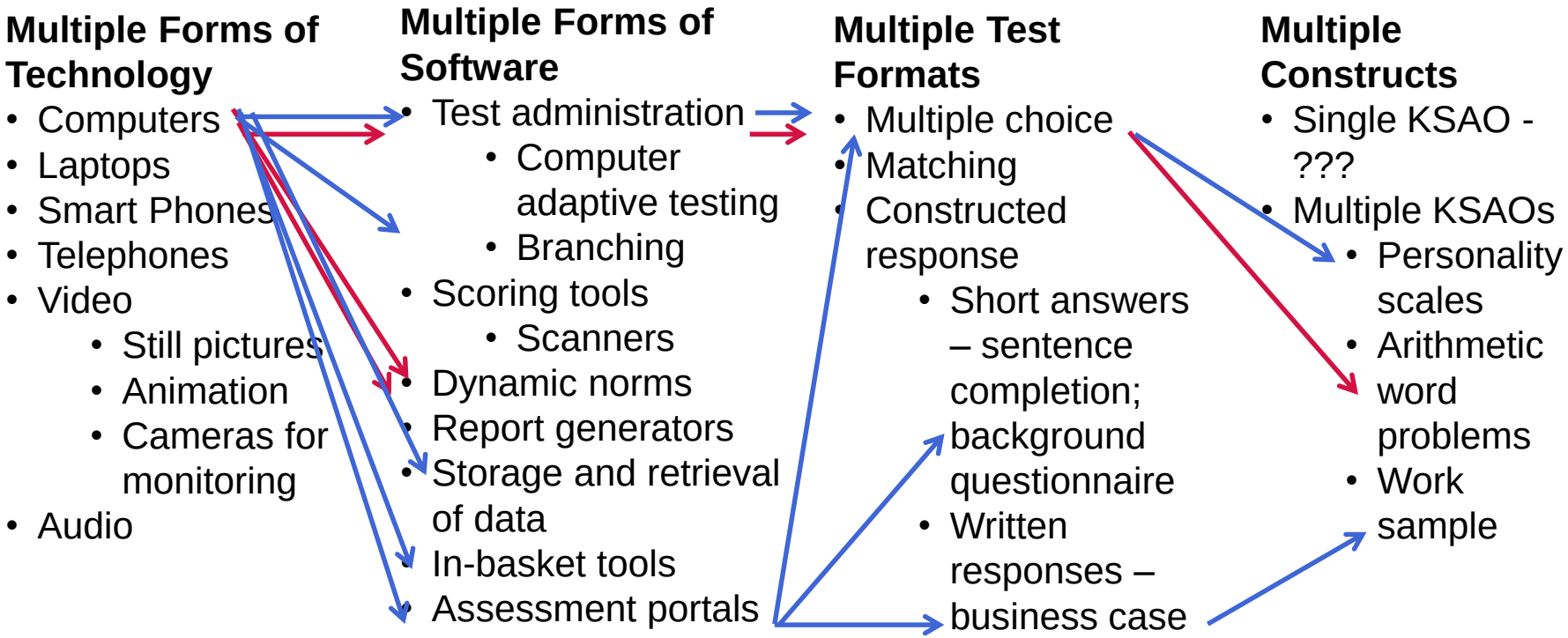
UIT

Validity
Cheating
Testing Environment
Applicant Pool
Equipment
Candidate Reactions
Score Equivalency
Costs
Ethical Responsibility

What is technology-enhanced assessment?

- Any assessment in which one or more component is facilitated by technology – administration, scoring, reporting, data storage

- A technology-enhanced assessment can be used in personnel assessment for selection, development, certification, evaluation



Advantages and Disadvantages

Advantages

- Costs
  - TAs
  - Printing
- Ease of updating
- Location of test – convenience
- Applicant reactions – simulations
- Candidate engagement – simulations
- Speed in processing
- Image/employment brand
- RJP – high fidelity

Disadvantages

- Costs
  - Software
- Potential for cheating and other forms of malfeasance
- Impact on validity
- Location of test – distractions
- Applicant reactions – CAT
- Privacy concerns
- Candidate engagement – virtual assessments

???

- Costs
  - Equipment
  - Internet connection
  - Bandwidth
- Applicant pool
- Diversity
- ADA accommodations
- Equivalence of scores
- Effect on adverse impact

What questions should testing professionals be asking?

- UIT
  - Is it valid? Is it valid enough? Do we have a responsibility to use the most valid test available to us?
  - Do people cheat? How much do they cheat? Who cheats? Does it matter?
  - What effect does the testing environment have on test taker performance? Who is responsible for ensuring a proper testing environment?
  - How does UIT affect the applicant pool in terms of size and diversity? Do technology-enhanced assessments really attract the right candidates?
  - Who is responsible for providing access to equipment, internet, and bandwidth?
  - What are candidate reactions? To what extent are candidates engaged?
  - To what extent are scores equivalent?
  - What are the costs? What is the cost/benefit ratio?
  - What does our Ethics Code require us to do?

Validity

- The validity of high fidelity, technology-enhanced work samples and simulations is almost unquestioned.
  - Many of the work samples and simulations are validated using a content-oriented strategy.

- In contrast, the validity of a UIT measuring cognitive ability is frequently questioned.
  - Does a UIT have sufficient validity to warrant its use in selection?
    - UITs have lower reliability; thus lower validity.
    - Individual studies
      - Validity of UIT is often less than the validity of a proctored exam.
      - Kaminski & Hemingway, 2009 found comparable validities for proctored and unproctored tests.
    - Meta-analysis
      - Beaty, Nye, Borneman, Kantrowitz, Drasgow, & Grauer, 2011 showed the validities of a proctored test and unproctored test to be comparable.

Validity

- Do we have a responsibility to use the most valid test available to us?
  - Many factors should be considered in selecting a test: validity, adverse impact, costs, resource requirements, time requirements, accessibility, compatibility with staffing process, etc.
  - We have a legal requirement to search for alternative selection procedures that have equal or higher validity and lower adverse impact.
    - The issue is probably not paper vs. computer testing; instead, it is more abstracted form of testing that may be computer-based vs. a high fidelity simulation that also incorporates technology (e.g., multiple-choice measure of critical thinking skills vs. computer-based in basket)

Cheating: Validity of an Individual Test Score

- How accurate is any one person's score?
- Does cheating occur?
  - There appears to be a small amount of cheating on unproctored tests.
  - Hense, Golden, and Burnett (2009)
    - Effect size of .32 between scores on a proctored and unproctored job simulation.
  - Arthur, Glaze, Villado, and Taylor (2009)
    - 7.7% of a sample cheated on speeded cognitive ability test (cheating defined as 1 SEM)
  - Lievens and Burke (2011)
    - Small d scores on cognitive test scores across four levels of jobs
      - Some in the opposite direction expected (proctored scores were higher than unproctored scores
      - <2.2% exhibited negative score change (only those in the high end of the distribution were retested)
    - Nye, Do, Drasgow, & Fine (2008)
      - No differences in scores from unproctored and proctored versions of an internet version of a perceptual speed test

Cheating

- Important considerations
  - Proctored testing often does not meet the gold standard for test administration (Drasgow, Nye, Guo, & Tay, 2009; Bartram, 2009; Foster, 2009)
    - Comparison may between some cheating and a lot of cheating
  - Cheating may not occur at equal rates across the entire distribution of test scores
    - Do smart people need to cheat?
  - Practice effect in second administration (usually the proctored test) minimize the amount of cheating observed
  - Higher motivation in the proctored setting (one of the later stages of the hiring process) may exist
  - Fewer distractions occur in the proctored setting, minimizing the extent of cheating observed
  - Better quality computer equipment may be used in the proctored setting
  - The real threat of verification testing may serve as a significant, effective deterrent to cheating

Cheating

- Final thoughts on cheating:
  - Deter or detect cheating: what is the better approach?
    - What are our best deterrents?
      - CAT
      - Verification testing
      - Warnings
  - How bad is it to hire someone who cheats on a test?
    - How much of a performance decrement will actually occur?
    - Is cheating on an employment test related to other dishonest behavior?
  - Is cheating on a cognitive ability test worse than distorting response on a personality inventory?
  - How do people cheat on work samples and simulations?
    - Pre-knowledge of materials
    - Outside help
    - Violation of time limits

Testing Environment

- Why does the test environment matter?
  - Validity and reliability of the test
  - Accuracy of the individual's test score
  - Applicant reactions
  - Size of the applicant pool

- Are our professional guidelines for test administration aligned with today's UIT practices?

Testing Environment

Standards for Educational and Psychological Testing

Standard 5.4

The testing environment should furnish reasonable comfort with minimal distractions.

Comment: Noise, disruption in the testing area, extremes of temperature, poor lighting, inadequate work space, illegible materials, and so forth are among the conditions that should be avoided in testing situations. The testing site should be readily accessible. Testing sessions should be monitored where appropriate to assist the test taker when a need arises and to maintain proper administrative procedures. In general, the testing conditions should be equivalent to those that prevailed when norms and other interpretative data were obtained.

Testing Environment

Standard 5.6

Reasonable efforts should be made to assure the integrity of test scores by eliminating opportunities for test takers to attain scores by fraudulent means.

Comment: In large-scale testing programs where the results may be viewed as having important consequences, efforts to assure score integrity should include, when appropriate and practicable, stipulating requirements for identification, constructing seating charts, assigning test takers to seats, requiring appropriate space between seats, and providing continuous monitoring of the testing process. Test developers should design test materials and procedures to minimize the possibility of cheating. Test administrators should note and report any significant instances of testing irregularity. A local change in the date or time of testing may offer an opportunity for fraud. In general, steps should be taken to minimize the possibility of breaches in test security. In any evaluation of work products (e.g., portfolios) steps should be taken to ensure that the product represents the candidate's own work, and that the amount and kind of assistance provided should be consistent with the intent of the assessment. Ancillary documentation, such as the date when the work was done, may be useful.

Testing Environment

Principles for the Validation and Use of Personnel Selection Procedures

Administration Environment - There are a number of factors that potentially affect test administration: appropriate workspace, adequate lighting, and a quiet, comfortable setting, free of distractions. The researcher should consider these conditions and their potential effects on test performance. At a minimum, selection procedure administration should be in an environment that is responsive to candidates' concerns about the selection procedures and maintains their dignity.

Testing Environment

- What effect does the testing environment have on test taker performance?
  - Weiner & Morrison (2009)
    - Testing environment is rated least favorably when test is administered in unproctored, on-site compared to on-site proctored and unproctored.
    - Non-cognitive scores are related to ratings of test environment.

- Who is responsible for ensuring a proper testing environment?
  - Should the employer act *in loco parentis* and dictate the test environment?
  - Should the employer provide the conditions under which most people do their best and allow the candidate to decide what conditions work best?

Applicant Pool

- How does UIT affect the applicant pool in terms of size and diversity?
  - UITs generally increase the size of the applicant pool
    - Applicant is not limited by time or location
  - Applicant pool may increase due to undesirable applicants
    - Some applicants are “window shopping”
    - Some lack the commitment to make further efforts related to employment
    - Some applicants are practicing
  - Could the applicant pool shrink because some candidates withdraw because of the practice of UIT?
  - Answers are difficult to come by because few comparative studies are done.

- How great is the digital divide?
  - In May, 2013, the Pew Research Center and American Family Life Project found
    - 91% of Americans own a cell phone
    - 34% a tablet
    - 61% and 58% own laptop or desktop computers

Applicant Pool

	% of adults who use the internet August 2011	% of adults who access the internet via broadband at home
All adults (age18+)	78	62
Men	80	65
Women	76	59
Race/ethnicity		
White, Non-Hispanic	80	66
Black, Non-Hispanic	71	49
Hispanic	68	51
Age		
18-29	94	76
30-49	87	70
50-64	74	60
65+	42	30

Pew Research Center and American Family Life Project

Size of Applicant Pool, Cheating, and Performance

- Landers & Sackett (2012)
  - Simulation study of effects of cheating and increases in size of applicant pool on performance
  - Conclusions:
    - If UIT increases the applicant pool, adopting UIT can result in higher mean criterion performance among those selected even if cheating occurs.
    - More cheating results in lower mean criterion scores, regardless of other factors.

Equipment, Internet Connection, and Bandwidth

- Other considerations regarding technology and its impact on testing:
  - Video-based assessments often require a large amount of bandwidth that may not be available to all candidates.
  - Smart phones often have slow or even no connectivity in certain places which disrupts the testing session.

- Limitations of equipment and connectivity may have an impact on who
  - Can access a test
  - Can complete a test

Mobile Testing

- Much has been made over testing on mobile devices, but is it really desired?

- HR Professionals believe (Fallaw & Kantrowitz, 2013)
  - 23% believe candidates want access to application forms and assessments via mobile devices
  - 43% would provide assessments via smartphone if the option existed
  - 41% indicated they would allow candidates to take assessments via smart phone if research showed the comparability of test scores from mobile devices with those on computers
  - 23% indicated that allowing candidates to take assessments on mobile devices was inappropriate

Mobile Devices

- Only a small percent of people take cognitive tests on mobile devices.
  - Hawkes (2013)
    - 1% of applicants take a graduate level cognitive test using a mobile device and 7% use a mobile device for a non-cognitive assessment.
  - Doverspike, Arthur, Taylor, & Carr (2012)
    - 1.7% of over 1M applicants used a mobile device to complete an assessment consisting of measures of verbal and numerical ability and a five-factor model measure
      - Caucasian (53.8%), female (59.0%) and under the age of 40 (74.3%)
      - Test takers using the mobile device performed less well than those using a PC; however, no differences were found on the non-cognitive measures.
  - Morelli, Illingworth, Scott, & Lance (2012)
    - Also found no differences when the assessment task involved a non-cognitive test.

Score Equivalency

- Cognitive ability
  - Mead and Drasgow (1993)
    - Equivalence between scores from paper-and-pencil and computer administered tests for power cognitive ability test
    - No equivalence between scores from each medium on speeded cognitive ability tests
    - No differences between adaptive and conventional tests administered via computer.
  - Potosky & Bobko (2004)
    - Modest degrees of cross-mode equivalence between scores from paper and pencil and internet based cognitive ability tests (timed).

Score Equivalency

- SJI
  - Potosky & Bobko (2004)
    - Modest degrees of cross-mode equivalence between scores from paper and pencil and internet
  - Ployhart, Weekley, Holtz, & Kemp (2003)
    - Variance-covariance matrices are not equivalent for scores from paper and web-based forms

- Biodata
  - Ployhart, Weekley, Holtz, & Kemp (2003)
    - Variance-covariance matrices are not equivalent for scores from paper and web-based forms

Equivalence

- Personality Test
  - Ployhart, Weekley, Holtz, & Kemp (2003)
    - Variance-covariance matrices are not equivalent for scores from paper and web-based forms measuring conscientiousness, agreeableness, and emotional stability
  - Mead, Michels, and Lautenschlager (2007)
    - Comparability of test scores on some (including conscientiousness) but not all personality constructs.
    - Lack of measurement invariance when those with choice of medium are compared to those without a choice.
  - Chuah, Drasgow, & Roberts (2006)
    - Found equivalence between traditional paper and pencil measures of personality and internet measures from unproctored testing sessions
  - Salgado, & Moscoso (2003)
    - Found equivalence of scores form paper and pencil measure of a Big Five personality measure to an internet based version

Equivalence

- Other thoughts on score equivalence:
  - Many organizations infer equivalency by comparing means and standard deviations, rather than establishing measurement invariance across constructs (Vandenberg & Lance, 2000)
    - Sample size is a significant problem
    - Comparisons across multiple devices and multiple cultures compound the difficult
  - Few studies compare high fidelity simulation to a paper-and-pencil test.
    - Implicit assumption is that these are measuring something different.
    - Chan & Schmitt (1997)
      - Compared written and video SJI – not an equivalence study; attributed different candidate reactions to the reading comprehension demand in the written form

Candidate Reactions

- Candidate reactions have long thought to have an impact on behavioral outcomes such as
  - Job offer acceptance (Boudreau & Rynes, 1985; Murphy, 1986) or become
  - Consumption of the organization's products and services (Rynes and Barber, 1990)
  - Reduction of the applicant pool(Boudreau & Rynes, 1985; Murphy, 1986)
  - Probability of a challenge to the selection procedures (Rynes, 1993; Rynes et al., 1980; Smither et al., 1993).

- Research does not confirm the relationships between applicant reactions and variety of outcomes
  - No relationship between applicant perceptions of the testing process and applicant withdrawal (Ryan et al., 2000; Truxillo et al., 2002; Hausknecht et al.'s, 2004)

Candidate Reactions

- Why?
  - Too few studies
  - Too few outcomes included
  - Global measures of candidate reactions
    - Ryan & Huth (2008) emphasize the need for specificity in evaluating candidate reactions to various components of an assessment

- Sackett and Lievens (2008) characterized this lack of evidence for a relationship between applicant reactions and individual or organizational outcomes as “the Achilles hell of this field” (p. 439).

Candidate Reactions

- Two findings illustrate the diversity of candidate reactions to technology-enhanced assessment:
  - Simulations and work samples generate more favorable reactions to simulations and work samples than to paper-and-pencil tests (Schmidt, Greenthal, Hunter, Berner, & Seaton,1977; Smither et al., 1993; Rynes & Connerly, 1993; Hausknecht, 2004)
  - Continuous monitoring via technology is perceived as more invasive and threatening to privacy; applicants may prefer one-time check such as biometric identification (Karim & Kaminsky; 2013)

Costs

- Few people report the costs of their technology-enhanced assessments.
- What are the sources of costs?
  - Development
    - Psychological assistance (e.g., validation research, item development and research)
    - IT Support (programming, maintenance of infrastructure, security activities)
    - Actors and videographers
  - Maintenance
  - Management
  - Administration
  - Equipment and internet connections
  - Scoring
  - Reporting
  - Per applicant fees (increase with the number of applicants)

Benefits

- More important question: What are the benefits that offset the costs?
- Few people report the benefits of their technology-enhanced assessment sin precise terms.
  - One exception is Cucina et al. who did not report costs or include estimates of SDy but did share utility estimates. The average utility for a video-based assessment tool (using only significant validities) was $109,183,823.
- What are the sources of benefits?
  - Reduced costs of administration (TAs, scorers, reporters)
  - Increased applicant pool
  - Better applicant pool
  - More efficient/timely staffing
  - More positive candidate reactions to high fidelity simulations (Schmidt, Greenthal, Hunter, Berner, & Seaton,1977; Smither et al., 1993; Rynes & Connerly, 1993; Hausknecht, 2004)
  - Additional data (e.g., response times)
  - Company image

Cost/Benefit

- Final thoughts:
  - Most organizations don't have a firm idea of the costs or the value of the benefits.
    - Even the tangible costs/benefits are difficult to evaluate; the intangible ones are almost impossible
  - Different factors have different weights for different organizations.
    - For example, some value the image they present to candidates more than others.

Ethics

- What does our Ethics Code require us to do?

(a) Psychologists administer, adapt, score, interpret, or use assessment techniques, interviews, tests, or instruments in a manner and for purposes that are appropriate in light of the research on or evidence of the usefulness and proper application of the techniques.

(b) Psychologists use assessment instruments whose validity and reliability have been established for use with members of the population tested. When such validity or reliability has not been established, psychologists describe the strengths and limitations of test results and interpretation.

(Section 9.02, Use of Assessments)

Ethics

Psychologists who develop tests and other assessment techniques use appropriate psychometric procedures and current scientific or professional knowledge for test design, standardization, validation, reduction or elimination of bias, and recommendations for use. (Section 9.05, Test Construction)

Psychologists who offer assessment or scoring services to other professionals accurately describe the purpose, norms, validity, reliability, and applications of the procedures and any special qualifications applicable to their use. (Section 9.09a, Test Scoring and Interpretation Services)

Psychologists make reasonable efforts to maintain the integrity and security of test materials and other assessment techniques consistent with law and contractual obligations, and in a manner that permits adherence to this Ethics Code. (Section 9.11, Maintaining Test Security)

Ethics

- Do we have an ethical problem?
  - It depends on what you did and what you say about you did.
    - Did you validate the UIT under the same conditions under which it will be used?
    - Are you disclosing the limitations of the UIT score?

Nancy T. Tippins
SVP
ntippins@executiveboard.com
+1 (864) 527-5956

1919 North Lynn Street
Arlington, VA 22209
telephone: +1 (571) 303-3000
facsimile: +1 (571) 303-3100