Data Integration in Psychological Assessment of Law Enforcement Personnel

Casey O. Stewart, Psy.D., ABPP
Corey & Stewart Consulting Psychologists
The Presentation

- Psychologists use a number of methods in conducting psychological evaluations of police candidates or incumbents (e.g., life history information, review of collateral informants/records, clinical interview, and psychological testing).
- Reaching a determination about suitability or fitness is straightforward when the data from these various sources all converge, but this is often not the case.
- Ultimately, it is the integration of discrepant data from multiple sources that poses the greatest challenge to examining psychologists.
- The presentation discusses the common methods of data collection in police candidate and incumbent psychological assessment.
- The presentation will also offer strategies for integrating data (e.g., assigning appropriate relative weight to and combining the different data) in a way to maximize the validity of suitability and fitness determinations.
Multiple Sources

• According to the Specialty Guidelines for Forensic Psychology (2011) “Forensic practitioners ordinarily avoid relying solely on one source of data ....”

• Both the Psychological Fitness-for-Duty Evaluation Guidelines (2009) and Pre-Employment Psychological Evaluation Guidelines (2009) indicate that evaluations should rely on multiple methods and data sources to optimize reliability and validity of findings.

• Obtaining information from a variety of sources is standard across high risk evaluations (Packer & Grisso, 2011).
Multiple Sources

• There are several reasons why such a wide variety of sources is considered best practice in high stakes assessment.

• All methods have inherent error, and the use of multiple methods to cross-check information reduces the potential for one’s conclusions to be based on error associated with any single information source (Packer & Grisso, 2011, p. 77)

• In order to increase the accuracy of clinical judgment clinicians “need to know how errors might occur, how to correct these errors, and the relative advantages of specialized training.” (Groth-Marnat, 2009, p. 24)
Multiple Sources

- According to Heilbrun, Grisso and Goldstein (2009) and Packer and Grisso (2011) the sources of data in FMHA include:
  1) self-report (face-to-face interview),
  2) one or more standardized psychological test,
  3) collateral informants (others who have had direct contact with the examinee),
  and 4) other relevant records (e.g., health care).
We Look for Convergence


• But data from the various sources do not always converge.

• How does this happen?
Garbage in Garbage Out (or insufficient info)

• According to the Specialty Guidelines for Forensic Psychology (2011), examiners “use assessment procedures in the manner and for the purpose that are appropriate in light of the research on or evidence of their usefulness and proper application” (EPPCC Standard 9.02, American Educational Research Association, American Psychological Association, and National Council on Measurement in Education [in press])
Conclusions About the Value of Different Sources of data

• Cognitive testing appears to be the most reliable and strongest predictor of job performance.
• “Biographical Data” appears to be second only to cognitive testing.
• Psychological testing has demonstrated moderate predictive and incremental validity.
• The clinical interview appears to be a valuable method, but is plagued with issues related to reliability that decrease it’s overall accuracy: The semi-structured interview appears to be the best option.
Incremental Validity

• According to Levy (2006), psychologists should 1) carefully select what predictors they use in their selection batteries, 2) realizing that using multiple predictors is likely to account for more variance in the criterion and that 3) some predictors tend to be more useful than others.
Incremental Validity

• Importantly, some predictors (e.g., clinical interview) may reduce the validity of the test battery (Aamodt, 2004).

• The examiner should use a management strategy that includes multiple sources of information (Melton et al., 2007).

• Indeed, important interpretations must be formulated with the benefit of other sources to protect against bias and reduce error (Packer & Grisso, 2011).

• Throughout the assessment the examiner “should integrate data and serve as an expert on human behavior rather than merely an interpreter of test scores.” (Groth-Marnat, 2009, p. 28)
Data Integration and Decision Rules

• According to Johnson (2011), forensic evaluations “weave together multiple relevant data sources as a by-product of forensically relevant judgments,” (p. 223)

• Cuttler (2011) argues that the use of decision rules is more accurate (i.e., includes less error) than “intervention by professional judgment,” (p. 157) for certain outcomes.
Data Integration and Decision Rules

• The assessment is most useful when it examines specific individual problems and provides guidelines for decision making regarding these problems (Groth-Marnat, 2009).
Data Integration and Decision Rules

• Although it is unlikely that actuarial prediction rules will replace clinical judgments, formal decision rules can and should be used more extensively as a resource to improve the accuracy of decision making” (Groth-Marnat, 2009, p. 28).

• “the process by which the data is aggregated” (Cuttler, 2011, p. 157) may be the most important scientific issue, in the area of assessment, that police psychology faces today.
Toward formal Decision Rules in Data Integration

- The accuracy of our decisions is a function of organization/structure and our adherence to procedural rules for decision making.
The Process

• Accuracy increases when interviewers are held accountable for the process they went through when coming to their decisions, compared to being held accountable for the accuracy of their predictions (procedural versus outcome accountability; Brtek & Motowidlo, 2002).
Data Analysis Procedures

• Most authorities on forensic reporting recommend that you list and describe all relevant data employed in the evaluation (Packer & Grisso, 2011).

• According to the Specialty Guidelines for Forensic Psychology (2011), “In their communication, forensic practitioners strive to distinguish observations, inferences, and conclusions.” (p. 15).
Data Analysis Procedures

- Example: W is what I conclude, because X is what I observed, which I believe to mean Y based on Z.
- I find Mr. Jones to be unsuitable for the position of police officer based on the testing that suggest 1) antisocial attitudes (So) and behaviors (RC4) as well as impulsiveness (Sc), which is empirically linked to negative job outcomes; 2) the background investigation (i.e., personal history data) evidencing a history of rule- and norm- violating behavior (e.g., failure to settle debt, non-felony offenses), linked in the literature to integrity violations; and 3) the interview, which revealed deceptiveness through inconsistencies, a claimed lack of memory around important life history questions and the response of “I don’t know”—all of which are the hallmark of deception.
- What if just two of these factors were present?
- What if the interview was negative (above), testing good (but defensive), and no admissions on the PHQ (which the agency uses for background)?
Data Analysis Procedures

• According to Packer and Grisso (2011), it is essential that all data relevant to the conclusions be included; this includes data that may weigh in the opposite direction (p. 98).

• You should explain how the data led to your specific conclusion, including, when relevant, an analysis of contradictory data and alternative explanations that were considered and the rationale for not accepting the alternative conclusion.

• Again, you should clearly convey the logic of the analysis tying the conclusions reached to the data reported (Packer & Grisso, 2011, p. 98).
Data Analysis and Integration

• The reliability or meaning of information from one source must be weighed in light of information obtained from other sources.

• One report may receive little weight in the examiners final opinion given information from other sources (Packer & Grisso, 2011, p. 77)
Integration

- According to Heilbrun, Grisso & Goldstein (2009), one of the most valuable ways psycho-legal assessments can be conducted is to “integrate ideographic evidence with nomothetic data.” (p. 112)

- Clinical judgments are more accurate when interview data is combined with formal assessments and when statistical interpretive rules were used.” (Groth-Mranat, 2009)

- Structured professional judgment involves the clinician determining how much weight to give various factors and then determining whether the individual is low-, moderate-, or high-risk.

- Indeed, Douglas, Ogloff, and Hart (2003) found that clinicians’ overall rating of risk (high, moderate, or low) were equal to, and in some cases superior to, numerical algorithms, supporting the validity of using professional judgment to weigh variables.
Utilize Actuarial Approach Whenever Possible

• It is unlikely that actuarial prediction rules will replace clinical judgments.

• But “formal prediction rules can and should be used more extensively as a resource to improve the accuracy of clinical decision making” (Groth-Marnat, 2009, p. 28).

• If actuarial formulas are available, they should be used when possible (Groth-Marnat, 2009).
Psychologist-Engineer

• Thomas (2010) saw psychologists as technologists and quoted Petroski (2003), who said that technologists solve problems, often using whatever science applies but equally often being ahead of science.

• According to Petroski (2003), “Engineers use science to solve their problems if the science is available. But available or not, the problem must be solved, and whatever form the solution takes under these conditions is called engineering.”

• The psychologist consultant is in many ways like the engineer. A given assignment may require approaches and solutions that the science of psychology has not yet addresses, or if addressed, not yet resolved (Thomas, 2010).
Absent a Calculus

• The future of psychological assessment will likely involve increasingly complex decision rules (actuarial tradition), which will likely be enhanced through computer technology.

• However, until a calculus is developed whereby all sources of data can be combined to produce a composite score predicting future outcomes, the process of data synthesis will be left to the psychologist.
Emulate Science

• That said, in the spirit of Heilbrun, Grisso and Goldstein (2009), Principles of Forensic Mental Health Assessment, we should model our clinical decisions after the best that science has to offer (actuarial tradition).

• We do this by creating models that contain decision rules that we apply mechanically.
Integration

• Indeed, according to Cuttler (2011) “... simple models can be built by consistently applying expert rules ...” (p. 157).

• The focus here is the process of combining the data, not what information you collect or how you collect it.

• Cuttler (2011) also suggests tracking outcomes and modifying the rules as indicated.
BOP Screening Enhancement Project

• The U.S. Department of Justice, Office of the Inspector General (OIG), conducted a reviewed to examine if the Federal Bureau of Prisons’ (BOP) hiring process could more effectively identify potentially unsuitable applicants for Correctional Officer positions.

• They used statistical procedures (classification tree, conditional inference tree, and logistic regression analyses) to identify applicant characteristics that correlate to future misconduct.

• The analyses show that combinations of certain applicant background characteristics are predictive of officer misconduct, but the BOPS’s hiring process does not have a systemic method of evaluating combinations.
BOP Screening Enhancement Project

• Educational level and duration of the longest civilian job predicted good behavior.
• Particular combinations of seven characteristics predicted a higher likelihood of substantial misconduct.
• Specific characteristics that make up the combinations matter less than the fact that combinations matter.
• Systematically considering combinations of characteristics and assigning weights to derive a risk rating can enhance screening accuracy.
Combinations of factors are a useful indicator of future behavior only when present in particular combinations

- Discipline at past jobs
- Unfavorable separation
- One job or less with supervisory responsibilities
- Having past due debt
- Having relatives that are inmates
- Using marijuana
- Working for less than 9.8 years at longest-held civilian job
Combinations of characteristics that increase the likelihood of officer misconduct

- Combination 1: Disciplined at past job + all credit accounts current + past use of marijuana + worked for less than 9.8 years at longest-held civilian job.
Combinations of characteristics that increase the likelihood of officer misconduct

- Combination 2: Not separated from past job under unfavorable circumstances + fewer than 2 past supervisory positions + all credit accounts are not current + worked for less than 9.8 years at longest-held civilian job.
Combinations of characteristics that increase the likelihood of officer misconduct

- Combination 3: Separated from past job under unfavorable circumstances + All credit accounts are *not* current + worked for less than 9.8 years at longest-held civilian job.
Combinations of characteristics that increase the likelihood of officer misconduct

• Some combinations are counterintuitive (BOP Combination 1 includes “all credit accounts current” as one of three variables), as are some predictors (positive relationship between traffic tickets and commendations).
Combining Variables

Most of the known combinations are within sources:

• BOP example of personal history information combinations
• Algorithms that constitute certain Risk Ratings on the CPI Suitability Snapshot
• MMPI patterns of scores as in the Goldberg, Husemann, or Gonder Indices
We Already Combine Variables

• Most psychologists already have familiarity with the factors related to problems on the job and combine them to derive selection decisions:
  • Non-felony criminal offenses
  • Elevations on MMPI-2-RF IPP
  • CPI PID > 20
  • Excellent presentation in the clinical interview
• However, we may not have predetermined relative weights for such factors.
The Problem

• Most of us do not have a mechanized process that allows for procedural accountability.
• That’s is, we do not have well-articulated predetermined decisions rules with regard to what we know (expertise) to rely on for combining data and making ultimate decisions.
• Therefore, our decisions are vulnerable not only to the error inherent in each predictor, but the many sources of natural bias that come with human judgments.
The Solution

• Establish which sources are the most reliable
• Assign weight to data relative to other source
• Have a model to guides your analysis of data
• Combine the data in a way that minimizes error and maximizes accuracy
• Do this using expert rules
• Apply them consistently
• Check your results
Note on Weighing Data

• If you are going to value one source over another, you should have a well-articulated rationale for devaluing the other source (e.g., lower performance on cognitive testing due to appraisal apprehension).

• In this example, if you are going to pass the examinee, you should have some additional evidence of nervousness in evaluative settings (e.g., reticence in the interview), but no evidence of past performance errors due to anxiety (freezing up) or significant indications in personality testing of deficits in insecurity/self-confidence.

• However, in a scenario where you cannot discover the source of the discrepancy, you should defer to the source with the least error.
Example

• Testing: Elevated RC4, JCP and DISC-r. Slightly elevated validity indicators.
• Background: History of juvenile delinquency and no evidence of contemporary counterproductive behavior at or away from work.
• Interview: Claims his days of norm- and rule-violating behaviors are far behind him. Reports prosocial attitudes, goals, and activities.
• Collaterals: Hyperactive trouble-maker in youth. Has matured a lot.
Example

• Testing: Very low Socialization (So) score on the CPI.
• Background: Gainfully employed. No recent legal problems, but shoplifting as a youth. Suspended from high school because of ditching class. No indication of substance abuse problems. Family history of criminal behavior.
• Interview: Seemingly well-adjusted, albeit reserved. Reports a chaotic home environment in childhood and strained relations with his parents.
• Collaterals: Unremarkable.
Example

• The case of elevated validity indicators.
• We already have an algorithm for that!
• And there may be other factors at play.
• Is it cultural or an indication of CWB?
• We know its linked to both.
Steps to Improve Clinical Judgment

Steps to improve accuracy of clinical judgment (i.e., reduce error) (Groth-Marnat, 2009, p. 25-26):

1. Structure your approach to interviewing in order to avoid missing important information. This is especial important when the stakes are high.
2. Consider/List evidence that does not support their hypothesis. This should reduce bias (e.g., confirmatory).
3. Attend to base rates when making predictions.
4. Seek feedback on accuracy and usefulness of judgments. Include rating forms to track your success.
5. Be knowledgeable about the person or group you are testing.
6. Continuing education about trends in clinical judgment should update evaluators on past and present knowledge and approaches.
Principles of Forensic Mental Health Assessment (Heilbrun, Grisso & Goldstein, 2009) adapted from (Heilbrun, 2001)

• According to Heilbrun, Grisso and Goldstein (2009), an established principle of forensic mental health assessments is:

• Select the most appropriate model to guide data gathering, interpretation, and communication:

• Morse (1978) and Grisso (2003), adapted for PPS, suggests a simple model of 1) the existence of psychological deficiencies or disorder, 2) functional requirements related to the job, 3) the connection between the deficiencies and/or disorder and the functional job-requirements.
Principles of Forensic Mental Health Assessment (Heilbrun, Grisso & Goldstein, 2009) adapted from (Heilbrun, 2001)

Identify relevant factors:

- POST Dimensions and Checklist:

  POST Psychological Screening Dimensions.pdf

  POST Dimensions Checklist.pdf
Principles of Forensic Mental Health Assessment (Heilbrun, Grisso & Goldstein, 2009) adapted from (Heilbrun, 2001)

• Use multiple sources of information: the examinee has some incentive to distort the accuracy of their self-report. Also, even without intending to be deceptive, any single informant may simply offer an inaccurate or biased perspective. The same may be said for any single psychological measure. Review the available background information and actively seek important missing information.
Principles of Forensic Mental Health Assessment (Heilbrun, Grisso & Goldstein, 2009) adapted from (Heilbrun, 2001)

• Use relevance and reliability (validity) as guides for seeking information and selecting data sources. Agreement across multiple sources of data (e.g., self-report, collateral records, psychological testing) make it more likely that the information is accurate.
Principles of Forensic Mental Health Assessment (Heilbrun, Grisso & Goldstein, 2009) adapted from (Heilbrun, 2001)

• Obtain relevant historical information: Among other important contributions, a history constructed from multiple sources can help the examiner gauge the accuracy of the evaluatee’s self-report, which can be particularly important for considering response style.
Principles of Forensic Mental Health Assessment (Heilbrun, Grisso & Goldstein, 2009) adapted from (Heilbrun, 2001)

- Assess personal characteristics (i.e., personal variables and functional capacities) in relevant (according to established criteria) and accurate (reliable and valid) ways: This is dependent upon a thoughtful approach to selecting methodologies (e.g., tests and specialized instruments) validated for the particular purpose and for this population.
Principles of Forensic Mental Health Assessment (Heilbrun, Grisso & Goldstein, 2009) adapted from (Heilbrun, 2001)

- Assess job-relevant behavior (the POST project, included “a statewide job analysis of the capabilities or ‘Competencies’ required to perform the job of the entry-level uniformed patrol officer, and the personality traits most closely related to these Competencies.” (POST, 2007, p. i)
Principles of Forensic Mental Health Assessment (Heilbrun, Grisso & Goldstein, 2009) adapted from (Heilbrun, 2001)

• Use third party information: “This particular principle emphasizes the importance of (a) using collateral documents and third party information in establishing a multi-source history, and (b) determining whether self-reported information is consistent with other sources and therefore more likely to be accurate.” (p.112)
Principles of Forensic Mental Health Assessment (Heilbrun, Grisso & Goldstein, 2009) adapted from (Heilbrun, 2001)

- Use case-specific (ideographic) evidence in assessing psychological factors and conditions, functional abilities, and causal connections: Individualized assessment.
- Use nomothetic evidence in assessing psychological factors and conditions, functional abilities, and causal connections: Group-based data. Compare test results of the examinee to others in the “known group” using the norms developed from such groups.
Principles of Forensic Mental Health Assessment (Heilbrun, Grisso & Goldstein, 2009) adapted from (Heilbrun, 2001)

- **Use scientific reasoning** in assessing causal connection between psychological factors and/or conditions and job-related requirements/competencies: The process of psycho-legal assessment can be comparable to science. When the results obtained from one source of information (e.g., interview or psychological testing) are treated as hypotheses to be verified through further information obtained from additional sources of information, it may be particularly helpful. When such hypotheses are accepted or rejected according to how well they account for the greatest amount of information with the simplest explanation, the scientific principle of parsimony is applied.
Principles of Forensic Mental Health Assessment (Heilbrun, Grisso & Goldstein, 2009) adapted from (Heilbrun, 2001)

- Control evaluator bias through continuing education, self-monitoring, and consultation.
# Simple Decision Matrix: A Heuristic

<table>
<thead>
<tr>
<th>Source</th>
<th>Combo 1</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Background</strong></td>
<td>O</td>
</tr>
<tr>
<td><strong>Testing</strong></td>
<td>-</td>
</tr>
<tr>
<td><strong>Interview</strong></td>
<td>+</td>
</tr>
<tr>
<td><strong>Decision</strong></td>
<td>-</td>
</tr>
</tbody>
</table>