

The New (2018) SIOP Principles: Content, History, Perspective

Joel P. Wiesen, Ph.D.
(post-conference version)
Contact: jw@jpwphd.com

2019 Annual IPAC Conference
Minneapolis, MN; 7/16/2019

Wiesen (2019) IPAC Conference

1

Outline of This Talk

- Changes to the Principles, 2003 to 2018
 - What's new in the 2018 *Principles*
 - What was omitted from the 2018 *Principles*
- Comparisons with the 2014 joint *Standards*

Wiesen (2019) IPAC Conference

4

Print and Audio Links

- PowerPoints (pre-conference posted)
- Audio recording (not yet posted)
- Track changes version of the 2003 SIOP *Principles* showing changes to yield 2018
- Proximity word search app (not yet posted)
- <http://jpwphd.com/ipac2019>

Wiesen (2019) IPAC Conference

2

Principles, 2003 to 2018

- Revision was incremental
- Table of Contents is largely the same
- Whole paragraphs are the same (verbatim)
- Small, meaty changes within paragraphs
- Citations to journal literature
- Some new paragraphs
- A few new topics

Wiesen (2019) IPAC Conference

5

Get the *Principles* on the Web

- Just search on title
- Link on my website:
<http://jpwphd.com/ipac2019>
- <https://www.apa.org/ed/accreditation/about/policies/personnel-selection-procedures.pdf>

Wiesen (2019) IPAC Conference

3

Track Changes Version

- Thanks for permission for track changes:
- Paul Sackett, Co-Chair of Ad Hoc Committee on the Revision
- Eden King, SIOP President
- Mo Wang, SIOP Publications Officer
- Cambridge University Press, Publisher
- *Permission granted for this presentation*

Wiesen (2019) IPAC Conference

6

Changes, Topic by Topic

- List topics
 - Six main topics plus glossary
 - No chapter numbers
- Highlight changes
- 51 pages, including the Forward

Introduction

- Statement of Purpose
 - Minor changes only
- Selection Procedures Defined
 - More cognizant of international and big data
 - Distinguishes between what tests measure and how they measure it
 - Omits polygraph and projective measures in list of examples

Pages/Topics

- 3 Introduction
- 5 Overview of the Validation Process
- 10 Sources of Validity Evidence
- 3 Generalizing Validity Evidence
- 2.5 Fairness and Bias
- 16 Operational Considerations in Personnel Decisions

Key to *Principles*' Headings

- Chapter headings: large, blue font, all caps
SOURCES OF VALIDITY EVIDENCE
- 1st subhead: initial caps, bold, left justified
Criterion-Related Evidence of Validity
- 2nd subhead: blue font, all caps, left justified
CRITERION DEVELOPMENT

Citations

- No numbered chapters
- Citations in this presentation are in these formats:
- [page/column/paragraph] for 2 column page
 - e.g., [5/2/1] = page 5, column 2, paragraph 1
- [page/paragraph] for one column page
 - e.g., [22/1] = page 22, paragraph 1

Overview of the Validation Process

- Sources of Evidence
- Planning the Validation Effort
- Analysis of Work

Overview of the Validation Process

- Changed “desirable” to “critical” here:
 - In addition, where contradictory evidence exists, comparisons of the weight of evidence supporting specific inferences to the weight of evidence opposing such inferences are **critical**. [5/1/3]
- Need to read new edition carefully
 - Even the parts that look familiar!

Wiesen (2019) IPAC Conference

13

Planning the Validation Effort

- Clarification re: SMEs
 - “... qualified for the tasks they are asked to perform and **knowledgeable about the information they are asked to contribute.**” [7/1/3]
 - “... qualified for the tasks they are asked to perform and **content they are asked to contribute.**” (2003) [9/last]

Wiesen (2019) IPAC Conference

16

Sources of Evidence

- Additions often clarify or provide cautions
 - “appropriate structural model ... (e.g., a confirmatory factor analysis model)” [5/23]
 - “... if the conceptual framework requires a more complex structure, overall consistency among items may not provide appropriate evidence of the internal structure of the test.” (2003) [6/4]

Wiesen (2019) IPAC Conference

14

Analysis of Work

- Some updates in terminology
 - KSAOs **or competencies** (e.g., pages 2, 3, 7, 8, 11, 12, 39)

Wiesen (2019) IPAC Conference

17

Sources of Evidence

- Some changes to terminology
 - “work performance” for “job performance” [9/3]
- 2018 version still uses both terms
- Work performance appeared in 2003 also

Wiesen (2019) IPAC Conference

15

Sources of Validity Evidence

- Evidence of Validity Based on Relationships With Measures of Other Variables
- Criterion-Related Evidence of Validity
- Evidence for Validity Based on Content
- Evidence of Validity Based on Internal Structure

Wiesen (2019) IPAC Conference

18

Evidence of Validity Based on Relationships With ...

- Terminology change
 - “work performance” for “job performance”

Wiesen (2019) IPAC Conference

19

Criterion-Related Evidence of Validity

- Added to Criterion Deficiency section
 - A common deficiency is only measuring work performance that is expected to relate to the predictor KSAPs measured [11/2/3]
 - Cf., Sackett, Shewach & Keiser, 2017
 - Found AC more valid than tests of g – meta-analysis of studies that employed both
 - “Ability tests ... predict narrower, more cognitively loaded criteria than ACs” [1435/2/1]

Wiesen (2019) IPAC Conference

22

Criterion-Related Evidence of Validity

- An expanded paragraph on statistical power
 - A validation study report should include power analyses [10/2/1]
 - Correcting for range restriction and unreliability increases the correlation coefficient but also its standard error [10/1/last par – 10/2]

Wiesen (2019) IPAC Conference

20

Criterion-Related Evidence of Validity

- Added 1½ paragraphs on criterion reliability [11/last par – 12/1]
- “... one should clearly describe the measurement design used and clarify what sources of error are reflected in the reported indices of reliability (e.g., rater-specific, item-specific, or occasion-specific errors).”

Wiesen (2019) IPAC Conference

23

Criterion-Related Evidence of Validity

- New sentences on predictive vs concurrent findings by test area
- Predictive and concurrent generally give same findings for cognitive tests
- Predictive and concurrent may NOT give same findings for non-cognitive tests [10/2/4]

Wiesen (2019) IPAC Conference

21

Criterion-Related Evidence of Validity

- New sub-section added (paragraph) on archival data as criteria [12/1/4]
- Describes some possible pitfalls in using archival data. Are the data:
 - aligned with the work analysis
 - free from contamination
 - acceptably reliable
- Quality of archival data often not apparent

Wiesen (2019) IPAC Conference

24

Criterion-Related Evidence of Validity

- New sub-section added (one paragraph) on Predictor deficiency [12/2/last par]
- “When judging ... [predictor] deficiency ... professional judgment ... takes into account both psychometric and practical considerations, including systematic bias against subgroups”

Criterion-Related Evidence of Validity

- Rewritten and expanded sub-section (two paragraphs): Choice of Participants
 - How to evaluate convenience samples
 - Demographics
 - Motivation
 - Ability
 - Experience
- [13/2/2]

Criterion-Related Evidence of Validity

- Added paragraph on algorithms used to make selection decisions (e.g., automated scoring of resumes [13/1/3])
- Mentions danger of capitalizing on chance
- Requires cross-validation prior to use
- Not introduce bias against “relevant subgroups”

Criterion-Related Evidence of Validity

- Recognizes practical limitations to comparing subsamples due to small N
 - “No matter how important a subsample may be to the testing professional, when it is too small, it cannot be statistically compared with other subsamples in an appropriate manner until additional data are available.”
- [13/2/3]

Criterion-Related Evidence of Validity

- Completely rewritten and expanded sub-section (two paragraphs): Predictor reliability
- Calls for clarifying “the sources of error that are reflected in the reported indices of reliability.” [13/2/1]

Criterion-Related Evidence of Validity

- Under Strength of the Predictor-Criterion Relationship, in discussing effect sizes and statistical significance of predictor-criterion relationship, added:
 - standard errors
 - confidence intervals
- [13/2/5]

Criterion-Related Evidence of Validity

- Does NOT take a stand on statistical tests
- "... the scientific literature is still evolving with regard to significance testing and estimates of uncertainty for validities, including those that have been corrected for statistical artifacts."
[13/2/5]

Criterion-Related Evidence of Validity

- In section on combining predictors, added 3.5 paragraphs [14/2/2-5]
- Added references on weighting
- Cautions about combining predictors with different variances and covariances

Criterion-Related Evidence of Validity

- Added a call for standard error or confidence interval for any point estimate for correlations [13/2/5]
- Web calculators, e.g.:
https://www.psych.org/stats/R/CI_correln1.html
For $r=.3$, $N=300$, 95% CI = 0.19 - .40
For $r=.3$, $N=50$, 95% CI = .02 - .53

Criterion-Related Evidence of Validity

- Discusses (mentions)
 - compensatory and non-compensatory combination of predictors
 - multiple cutoffs on individual predictors

Criterion-Related Evidence of Validity

- Added some useful references to section: Adjustments To Validity Estimates [14/1/2-4]

Criterion-Related Evidence of Validity

- More on adverse impact
- Directs us to consider impact of weighting and cutoffs on subgroups [14/2/4]
- Mentions tradeoffs between validity and subgroup differences without proscribing such tradeoffs [14/2/5]

Criterion-Related Evidence of Validity

- Omitted a sentence favoring unit weights:
- “Usually, it is better to assign unit or equal weights to the several criterion components than to attempt to develop precise empirical weights.” (2003) [20/3]

Wiesen (2019) IPAC Conference

37

Evidence for Validity Based on Content

- Are psychologists SMEs?
- “SMEs ... establish the relationship between the selection procedures and the work behaviors or worker requirements.” [15/2/5]
- Problem: Incumbents do not understand Fleishman areas!

Wiesen (2019) IPAC Conference

40

Criterion-Related Evidence of Validity

- Cites in addition to Schmidt and Hunter (1998)
- “A large body of research regarding relationships between many predictors and work performance currently exists...” [15/1/3]

Wiesen (2019) IPAC Conference

38

Evidence for Validity Based on Content

- Added to content-based validity study
- “... other steps in this process ... **collecting SME judgments about the link between the selection procedure and the requirements of the job ...**” [16/1/2]

Wiesen (2019) IPAC Conference

41

Evidence for Validity Based on Content

- Introduces a new term:
- “... these selection procedures are labeled ‘**content-based predictors**’.” [15/2/2]

Wiesen (2019) IPAC Conference

39

Evidence for Validity Based on Content

- Repeats 2003 statement domain need not be exhaustive
- “The domain need not include everything that is done on the job. The testing professional should ... explain why certain parts of the domain were or were not included in the selection procedure.” [16/1/3]

Wiesen (2019) IPAC Conference

42

Evidence for Validity Based on Content

- Specifically allows content validation of personality characteristics, adding a phrase
- “The fact that the construct assessed by a selection procedure is labeled an **ability or personality characteristic** does not *per se* preclude the reliance on a content-oriented strategy.” [16/1/4]

Wiesen (2019) IPAC Conference

43

Evidence for Validity Based on Content

- New paragraph on Competency Modeling
- Calls for “A rigorous competency modeling study”
- Gives a citation for best practices [17/1/3]

Wiesen (2019) IPAC Conference

46

Evidence for Validity Based on Content

- Same as in 2003
- “The selection procedure should reflect **adequate coverage** of work behaviors and activities and/or worker requirements from this restricted domain to provide sufficient evidence to support the validity of the inference.” [16/2/3]

Wiesen (2019) IPAC Conference

44

Evidence for Validity Based on Content

- “... reliability estimate reported should reflect ... the measurement design ... [of the] selection procedure, the generalizations one wishes to make regarding the ... scores, and how the predictor measure will be used (e.g., for rank ordering applicants, or for making pass–fail or hire–no hire decisions; cf. Predictor reliability).” [17/2/4]

Wiesen (2019) IPAC Conference

47

Evidence for Validity Based on Content

- Added a phrase, requiring a test plan
- “The rationale underlying the sampling should be **documented in a test plan specifying which KSAOs are to be measured by which assessment methods.**” [16/2/last par]

Wiesen (2019) IPAC Conference

45

Generalizing Validity Evidence

- Transportability
- Synthetic Validity/Job Component Validity
- Meta-Analysis

Wiesen (2019) IPAC Conference

48

Transportability

- Elaborates on definition
- Cites to literature [19/2]

Meta-Analysis

- Meta-analysis has some small changes, most additions
- Updated citations to the literature

Synthetic Validity/Job Component Validity

- Provides some refs for synthetic validity [20/1/3,4]

Meta-Analysis

- Added
- Not all meta-analyses agree
- Not all are of high quality [20/2/last par – 21/1/1]

Meta-Analysis

- Omitted phrase from 2003 edition
- “While transportability and synthetic validity/job component validity efforts may be based on an original study or studies that establish the validity of inferences based on scores from the selection procedure through a **content-based** and/or a criterion-related strategy” (2003 *Principles*) [28/3]

Fairness and Bias

- Fairness
- Bias

Fairness

- Keeps various views, as in 2003
- Equal outcomes - *Standards* rejects this
- Equitable treatment in the selection process
- Comparable access to the constructs
- Lack of bias (incl. measurement bias) [22]

Wiesen (2019) IPAC Conference

55

Fairness

- New:
- “Most organizations strive for a diverse and inclusive workforce and equitable treatment of cultural and linguistic minorities.” [23/1/2]
- Nothing on acceptable/recommended methods
 - Mentions investigating causes for differences

Wiesen (2019) IPAC Conference

58

Fairness

- Under equitable treatment, mentions new modes of test administration
 - e.g., mobile devices
 - does not elaborate
- [22/3]

Wiesen (2019) IPAC Conference

56

Bias

- Keeps two types of bias of 2003 edition
 - Measurement bias
 - Irrelevant sources of variance in predictor
 - Predictive bias
 - Effects of irrelevant sources of variance on predictor-criterion relationships
- [23]

Wiesen (2019) IPAC Conference

59

Fairness

- Under comparable access to the constructs
 - Was: “**opportunity to learn**”
 - Now “comparable access to the constructs”
- [22/4]
- Now ends with: “restrict accessibility and affect measurement”

Wiesen (2019) IPAC Conference

57

Bias

- Expands on ways to test for predictive bias [23/1/last par]
- Adds literature summary on overprediction [23/2 – 24/1]

Wiesen (2019) IPAC Conference

60

Bias

- Added to measurement bias section
- Item sensitivity reviews
- DIF analysis (not expected or required)
[24/1]

Wiesen (2019) IPAC Conference

61

Principles on Fairness

- Keeps various views, as in 2003
- Equal outcomes - *Standards* rejects this
- Equitable treatment in the selection process
- Comparable access to the constructs
- Lack of bias (incl. measurement bias)
[22]

Wiesen (2019) IPAC Conference

64

Comparison with Joint *Standards*

- Main focus of this comparison is fairness
- Start by comparing some details
- Then generalize to a more global comparison

Wiesen (2019) IPAC Conference

62

Term Fair Defined Differently

- A profession usually has its own vocabulary
- Strange to have one profession with two very different definitions of fairness

Wiesen (2019) IPAC Conference

65

Definition of Fairness

- *Principles*: “Fairness is a social ... concept”
[22/1]
- *Standards*: Standard 3.0 on Fairness
“All steps in the testing process ... to minimize construct-irrelevant variance and to promote valid score interpretations...”
[63/1]

Wiesen (2019) IPAC Conference

63

Standards on Fairness

- “The central idea of fairness in testing is to identify and remove construct-irrelevant barriers to maximal performance for any examinee.”
[63/1, last par]

Wiesen (2019) IPAC Conference

66

Standards on Fairness

- Broad Goal of Testing
- “Achieving equality of opportunity in our society”
- “Fairness is a fundamental validity issue.”
AERA, APA, NCME (2014) [49/1/1, 2]

Wiesen (2019) IPAC Conference

67

Standards on Fairness

- “... fairness is a fundamental issue for valid test score interpretation, and it should therefore be **the goal** for all testing applications.”
AERA, APA, NCME (2014) [62/2/4]

Wiesen (2019) IPAC Conference

70

Standards on Fairness

- “... if differences were due to the test’s sensitivity to some test-taker characteristic not intended to be [measured], then the ... interpretation ... as predicting job performance in a comparable manner for all groups ... **would be rendered invalid, even if test scores correlated positively** with some measure of job performance.”
[21/1/1]

Wiesen (2019) IPAC Conference

68

Standards on Universal Design

- “By using universal design, test developers begin the test development process with an eye toward maximizing fairness.” [57/2/3]
- Fairness is of primary concern at every step in test development and use.

Wiesen (2019) IPAC Conference

71

Standards on Fairness

- “... fairness to all individuals in the intended population of test takers is an **overriding, foundational concern**”
AERA, APA, NCME (2014) [49/2/1]

Wiesen (2019) IPAC Conference

69

Joint Standards

- Aspirational goals vs standards
- *Standards* not superseded by *Principles*
- *Principles* does not restate all *Standards*
- Much agreement
- Some disagreement

Wiesen (2019) IPAC Conference

72

Standards vs Principles

- *Standards*' Universal Design more prominent than *Principles*' lack of bias
- *Principles*: has citations to the literature
- *Standards*: no citations to the literature

Standards vs Principles

- *Standards* gives examples of topics to consider in a follow-up study
- Construct underrepresentation
- Construct irrelevant variance
- Internal structure of test responses
- Response processes used by individuals [65/2/last par]

Standards vs Principles

- Different psychometric cautions
 - e.g., *Standards* [66/2/1], *Principles* [24/1/1,2]
- *Standards* mentions criteria bias elsewhere
- *Standards* mentions reliability by subgroup
- *Principles* mentions bias in criteria
- *Standards* mentions test security here
- *Principles* test security under data collection

Standards vs Principles

- *Standards*: 24 pages on fairness
 - 14,900 words, 2,400 sentences, w/ 20 Standards
- *Principles*: 2.5 pages on fairness and bias
 - 1,950 words, 260 sentences
- **The number of words/pages is misleading**
- Both give unique, substantive guidance
- Need to read both!

Standards vs Principles

- *Standards*: mean score differences trigger follow-up studies, where feasible
- *Principles*: ... predictive bias analysis should be undertaken ... compelling reasons ... to question whether a predictor and a criterion are related [comparably] ... for specific subgroups ... generalized evidence can be appropriate for examining predictive bias... [24/1/2]

2018 Principles

- Last "chapter":
- Operational Considerations in Personnel Selection
 - Development and choice decisions
- 15 pages (of a total of 40 pages)
- 14 main sub-headings

Operational Considerations in Personnel Decisions

- Initiating a Validation Effort
- Understanding Work and Worker Requirements
- Selecting Assessment Procedures for the Validation Effort
- Selecting the Validation Strategy
- Selecting Criterion Measures

Wiesen (2019) IPAC Conference

79

Initiating a Validation Effort

- New: Cost and cost-benefit analysis
– [26/1/2]

Wiesen (2019) IPAC Conference

82

Operational Considerations in Personnel Decisions (continued)

- Data Collection
- Data Analyses
- Appropriate Use of Selection Procedures
- Recommendations
- Technical Validation Report

Wiesen (2019) IPAC Conference

80

Initiating a Validation Effort

- New: Mentions “synthetic validation” in the list of validation strategies [26/1/5]

Wiesen (2019) IPAC Conference

83

Operational Considerations in Personnel Decisions (continued)

- Administration Information
- Other Circumstances Regarding the Validation Effort and Use of Selection Procedure
- Assessing Candidates With Disabilities
- Candidate Linguistic and Cultural Background

Wiesen (2019) IPAC Conference

81

Initiating a Validation Effort

- New: “A new work analysis should be conducted when test developers or users have reason to believe that the nature of the work performed has changed meaningfully since any prior analysis was conducted.”
[27/1/3]

Wiesen (2019) IPAC Conference

84

Understanding Work and Worker Requirements

- New: Paragraph on media for test administration:
- Proctored vs unproctored
- Video vs written
[28/1/3]

Wiesen (2019) IPAC Conference

85

Data Analyses

- More on missing data and outliers
- Missing data: “two commonly recommended strategies are full information maximum likelihood (FIML) and multiple imputation (MI) approaches”
- Outliers: “should also check their data for both univariate and multivariate outliers”
[31/1/last 2 pars]

Wiesen (2019) IPAC Conference

88

Selecting Criterion Measures

- New: “When reporting criterion reliability ... describe the type of reliability estimate and sources of error that are reflected in (and ignored by) the reliability index.”
[29/2/last par]

Wiesen (2019) IPAC Conference

86

Data Analyses

- New: “... non-independence (clustering of individuals) in the predictor–criterion data being analyzed can affect the accuracy/quality of inferences and should be considered.” [31/2/3]

Wiesen (2019) IPAC Conference

89

Data Analyses

- New: “... should consider including content or mechanisms to help identify careless or insufficient effort responding” [30/2/last par]

Wiesen (2019) IPAC Conference

87

Data Analysis

- New: “... some organizations may put more emphasis on maximizing validity ... other organizations ... more emphasis on minimizing subgroup differences relative to maximizing ...” [32/1/1]
- Legitimizes making selection system design decisions based on subgroup differences!

Wiesen (2019) IPAC Conference

90

Data Analysis

- More on evaluating criteria
- Should report the following in detail:
 - Description of ... measures
 - Rationale for their use
 - Data collection procedures

A Parting Gift

- Track Changes comparison, 2003 vs 2018
- On my website:
<http://jpwphd.com/ipac2019>

Data Analysis

- More on evaluating criteria
- Should report the following in detail:
- "... discussion of the measures' relevance, reliability, possible deficiencies, possible sources of contamination, and freedom from or control of biasing sources of variance."
[34/2/2]

References

- Some 140+ references
- Many new citations