



Evolution of Evaluation Checklists From Creation To Validation

Wes Martz, Ph.D.

Daniela C. Schröter, Ph.D.



Overview of Session

- 1** Introduction to Evaluation Checklists
- 2** OEC Development and Validation
- 3** SEC Development and Validation
- 4** Lessons Learned



Evaluation Checklists

- Mnemonic device
- Inherently systematic
- Provides guidance for collection of evidence
- Highly relevant for evaluative purposes



Taxonomy of Checklists

TYPE	EXAMPLE
Laundry list	Grocery lists, travel lists, party planner list
Strongly/weakly sequential	Flight crew checklist, product assembly checklist
Iterative	Medical procedures lists Several evaluation checklists (e.g., KEC, OEC, SEC)
Diagnostic	Psychological assessments (e.g., Diagnostic and Statistical Manual)
COMlist	Rating sheets in contests, Personnel selection qualities Sections in evaluation checklists (e.g., KEC, OEC, SEC)



Why develop checklists?

Advantages

- Consolidate vast knowledge in a parsimonious manner¹
- Improve task performance²
- Reduce influence of halo and Rorschach effects¹
- Reduce resource use³

Disadvantages

- Overuse (fatigue)
- Unnecessary barriers
- Tunnel vision
- Inappropriate use

¹ Scriven, 2007; ² Hale et al., 2007; Wolff et al., 2004; ³ Persaud, 2007



Validating Checklists

- Limited guidance available for validating evaluation checklists
- Valid checklists are important to validity and credibility of evaluative conclusions





Validation Methodology

ORGANIZATIONAL EFFECTIVENESS EVALUATION CHECKLIST (OEC)



OEC Overview

- Organizational evaluation process framework
- Iterative, explicit, weakly sequential
- 29 checkpoints grouped into six steps
- Criteria of merit checklist
- 12 universal criteria of merit
- 84 suggested measures

www.EvaluativeOrganization.com



OEC Validation Process

- **Phase 1: Expert panel review**
 - Critical feedback survey
 - Written comments made on checklist
- **Phase 2: Field test**
 - Single-case study
 - Semi-structured interview



OEC Validation Study

PHASE 1: EXPERT PANEL REVIEW





Expert Panel Overview

- Study participants
 - Subject matter experts (organizational and evaluation theorists)
 - Targeted users (professional evaluators, organizational consultants, managers)
- Review OEC for providing critical feedback
- Identify strengths and weaknesses
- Complete the critical feedback survey
- Write comments directly on the checklist



OEC Validation Process

Performance Criteria

- Applicability to full range of intended uses
- Clarity
- Comprehensiveness
- Concreteness
- Ease of use
- Fairness
- Parsimony
- Pertinence to the content area



Expert Panel Data Analyses

- Critical feedback survey
 - Descriptive statistics
 - Parametric and nonparametric analysis of variance
 - Net strength score
- Written comments on checklist
 - Hermeneutic interpretation
 - Thematic analysis to cluster and impose meaning
 - Triangulation across participants to corroborate or falsify the imposed categories



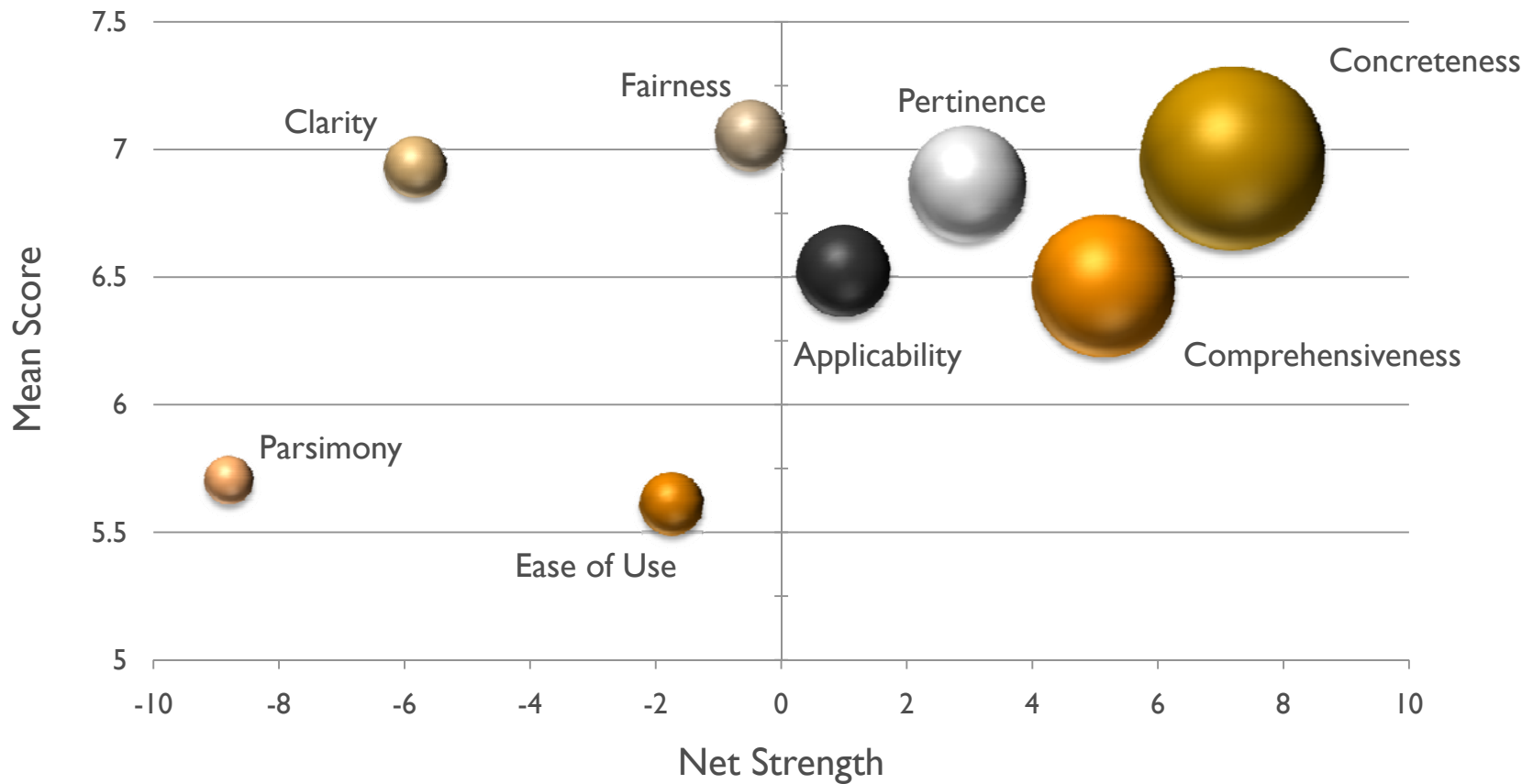
Net Strength Score

Net strength is the difference between the normalized frequencies of the strengths and weaknesses from the narrative critique.

	Strength		Weakness		Net Strength (S' - W')
	<i>Frequency</i>	<i>Normalized</i>	<i>Frequency</i>	<i>Normalized</i>	
	(S)	(S')	(W)	(W')	
Applicability	6	8	2	7	1
Clarity	13	18	7	24	-6
Comprehensiveness	11	15	3	10	5
Conciseness	10	14	2	7	7
Ease of use	11	15	5	17	-2
Fairness	7	10	3	10	0
Parsimony	6	8	5	17	-9
Pertinence	7	10	2	7	3



Net Strength Score vs. Mean





OEC Validation Study

PHASE 2: FIELD TEST





Field Test Overview

- Evaluation Client
 - For-profit organization based in the U.S.
 - Specializes in storytelling using electronic media
 - Stand-alone entity (e.g., not a division or program)
- Conduct evaluation using revised OEC
- Strictly follow OEC to ensure fidelity
- Post-evaluation semi-structured client interview
- A formative metaevaluation to detect and correct deficiencies in the process



Observations from Field Test

- Structured format minimized “scope-creep”
- Identified several areas to clarify in OEC
- Reinforced need for multiple measures, transparency in standards
- Minimal disruption to the organization
- Evaluative conclusions were credible, valid, and useful
- High value placed on recommendations



Aspects of OEC Validity

TYPE	EVIDENCE PROVIDED BY
Content	Professional judgments made by expert panel members
Substantive	Expert panel review and field study
Generalizability	Single-case field study, analytic generalizability
External	Comparison to other OE assessment models
Consequential	Expert panel review (interpretation) and field study (use)



OEC Validation Method

Strengths & Weaknesses

- Strengths
 - Relatively quick validation process
 - Based on relevant evaluative criteria
 - Features a real-world application
- Weaknesses
 - Single-case field study
 - Selection of the case study
 - Selection of the expert panel members



Validation Methodology

SUSTAINABILITY EVALUATION CHECKLIST (SEC)



Overview of the SEC

Purpose & Users

- Purpose: Planning and designing project and program evaluations *of sustainability for sustainability* in development contexts
 - Remember evaluation tasks
 - Consider criteria of sustainability
 - Reduce errors of omission
 - Increase evaluation usefulness
 - Generate ideas and discussion among stakeholders
 - Support proposal writing
 - Think about existing sustainability evaluations
- Users:
 - Evaluators and researchers
 - Program planners, funders, managers
 - Recipients or participants



Overview of the SEC

Characteristics

- The SEC is:
 - A mnemonic device
 - Weakly sequential
 - A “meta-tool” intended to help planning and designing sustainability evaluations
- The SEC is not:
 - A survey instrument
 - Measurement device
- The SEC requires:
 - Consideration and iteration
 - Critical thinking



SEC Overview

www.SustainabilityEval.net



Introduction:
Purpose, Intended Users, Characteristics, Key Concepts



Part A - General Considerations in Evaluation:
Grounding the Evaluation, About the Evaluand, General Procedures for Evaluation



Part B: Criteria in Sustainability Evaluation:
Criteria of Merit, Criteria of Worth, Criteria of Significance



Appendices:
Glossary, References



SEC Development & Validation

- Development
 - Literature-based development imbued with iterative feedback from experts and practitioners
 - Continuous process
- Validation
 - Nonexperimental, exploratory, mixed method study
 - Systematic feedback on checklist accuracy and utility via
 - Self-administered Web-based questionnaire (N=106)
 - Expert interviews (N=10)
 - Email (N=5+)



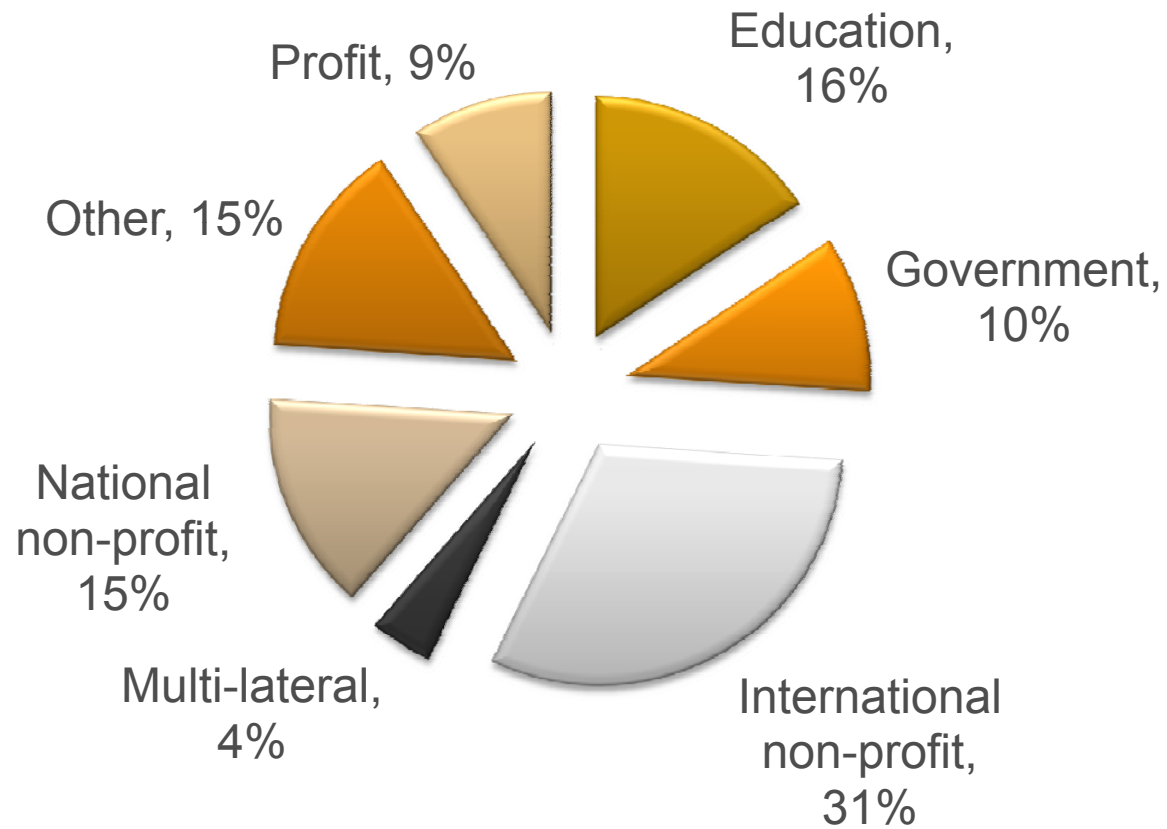
SEC Study

Participants

- 111 pre-identified experts and practitioners in
 - Evaluation
 - Sustainable development
 - International development
- Sampling
 - Purposeful maximum heterogeneity sampling
 - Expert recruitment via direct contact
 - Practitioner recruitment via Listservs

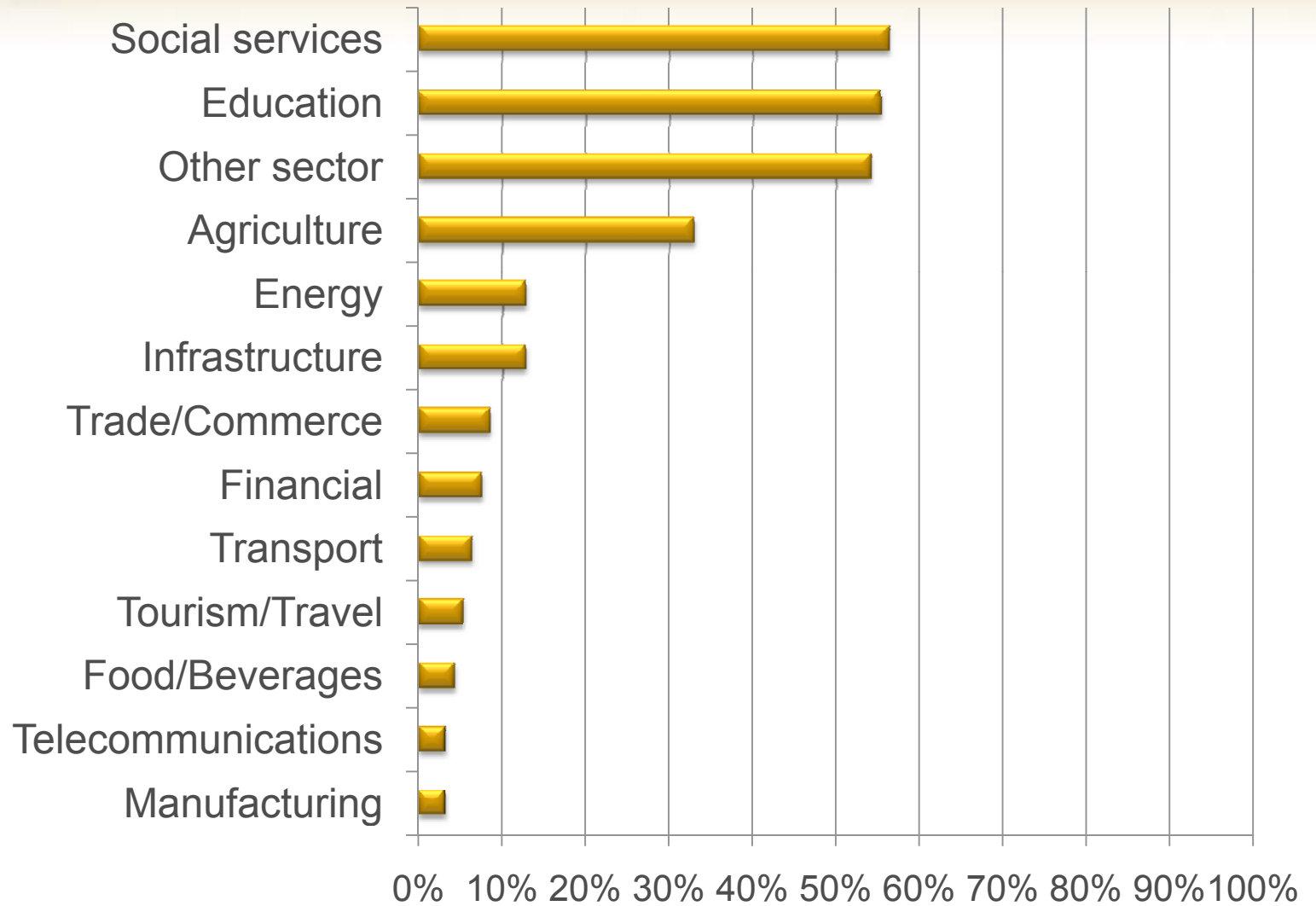


Participants: Organizational Affiliation



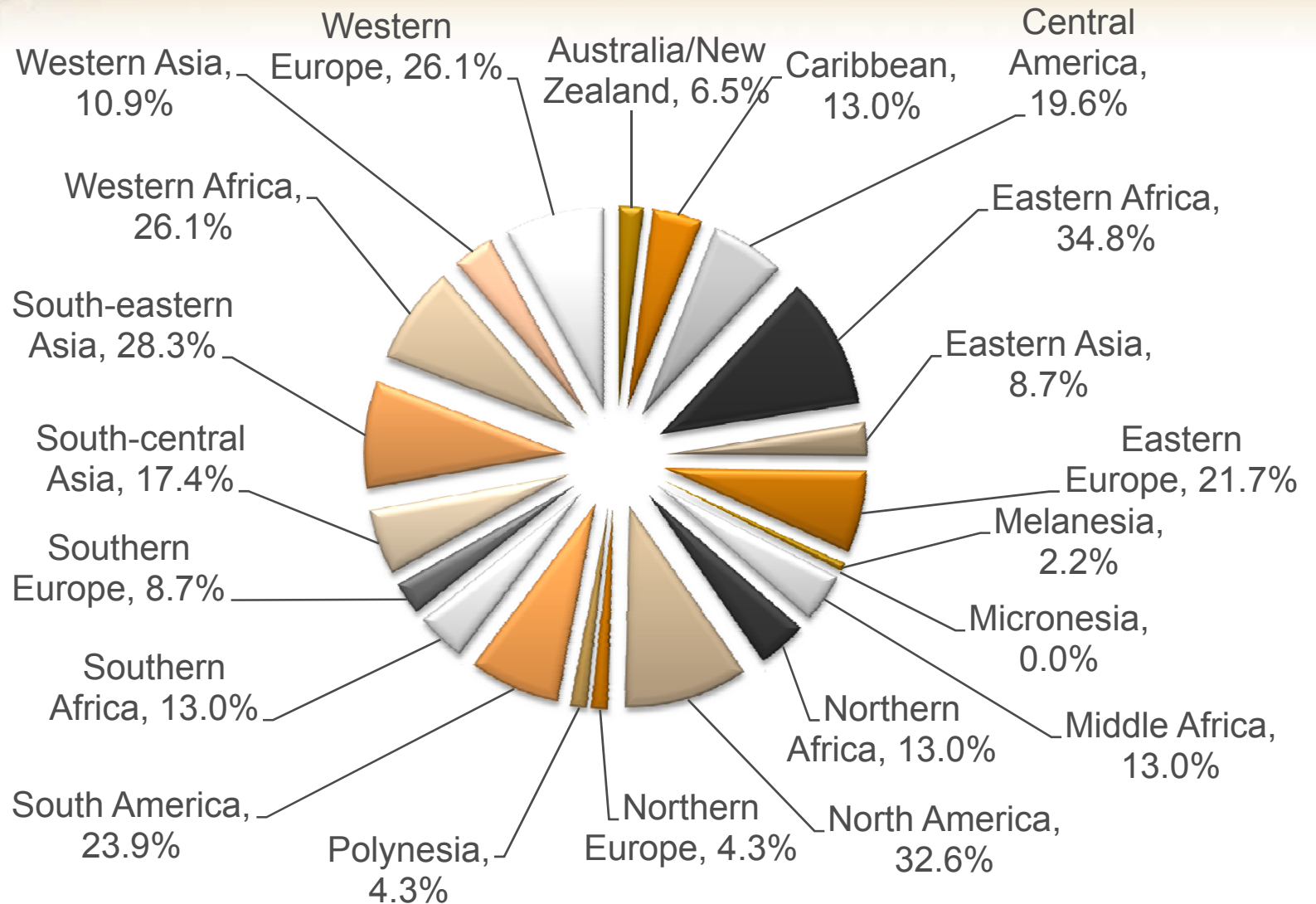


Participants: Sectors

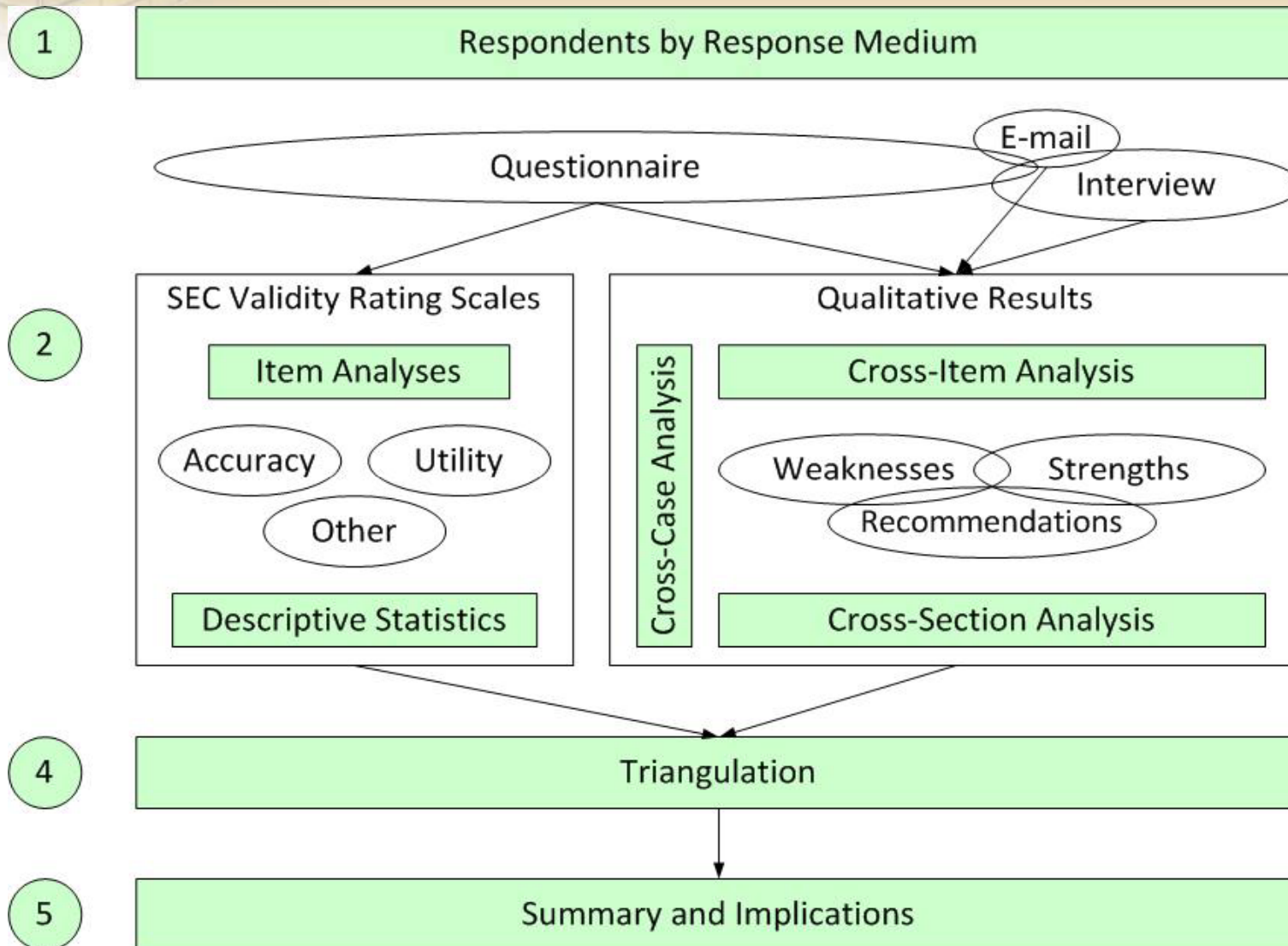




Participants: Work Regions



SEC Validity Study: Analytic Framework





SEC Validity Measures

- Measures
 - 25 items measuring checklist accuracy and utility
 - Five point agreement scale from 0 – not at all to 4 - completely
- Item analyses
 - 7 items were removed
 - Convergent and discriminant validity of remaining items
 - Internal consistency estimates/reliability
 - Accuracy scale: $\alpha = .87$
 - Utility scale: $\alpha = .89$

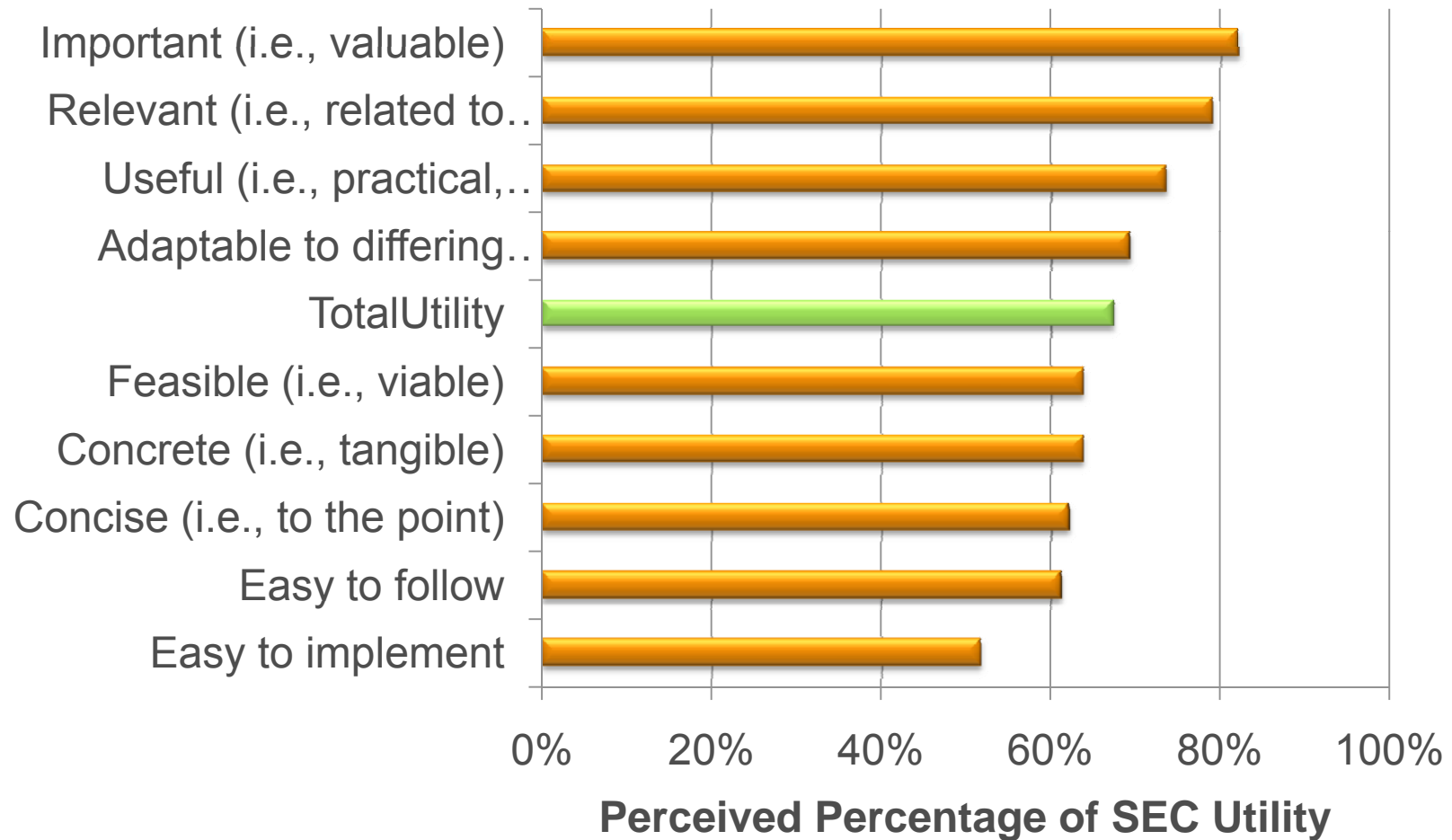


SEC Accuracy



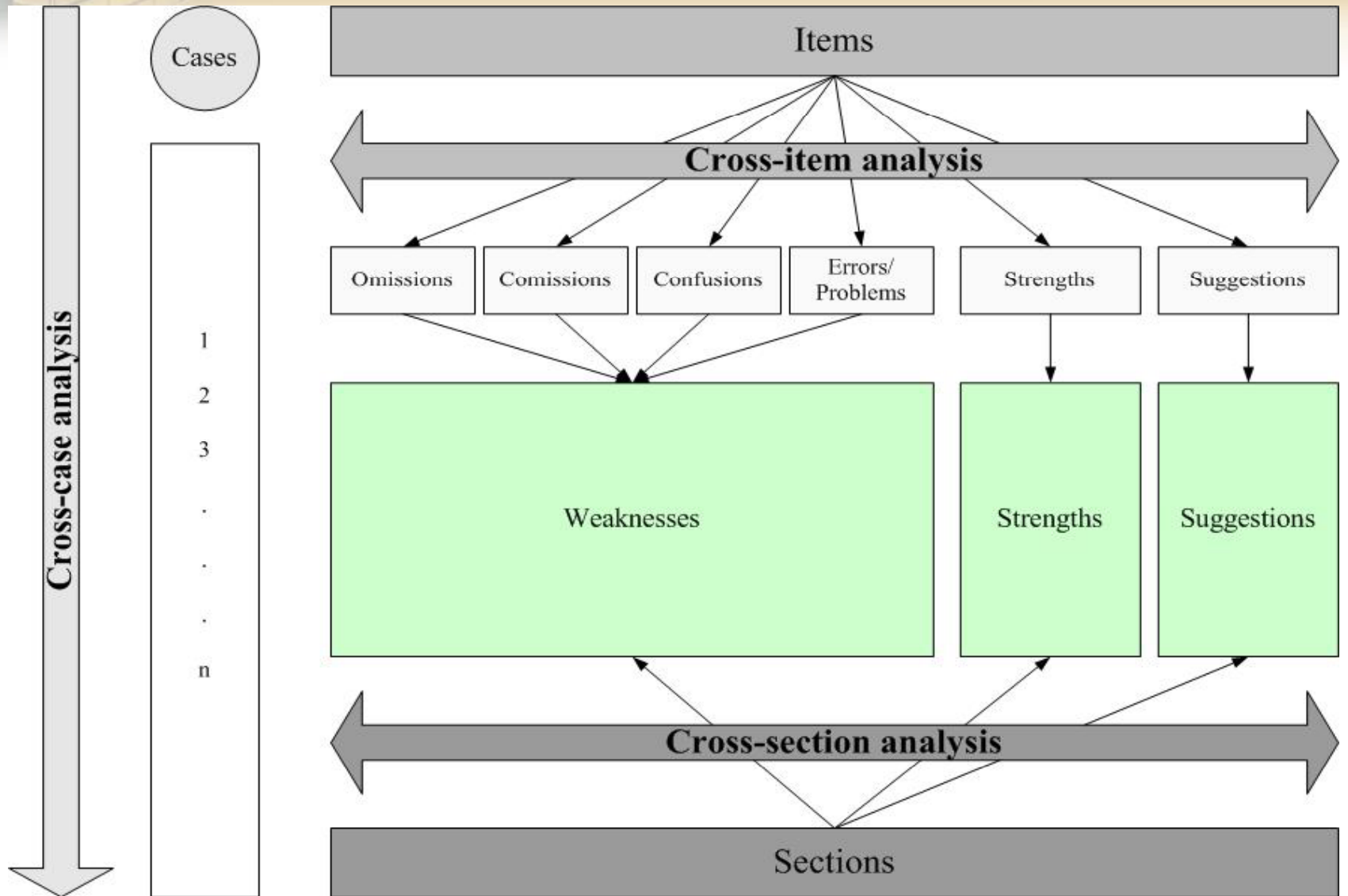


SEC Utility





Qualitative Analysis



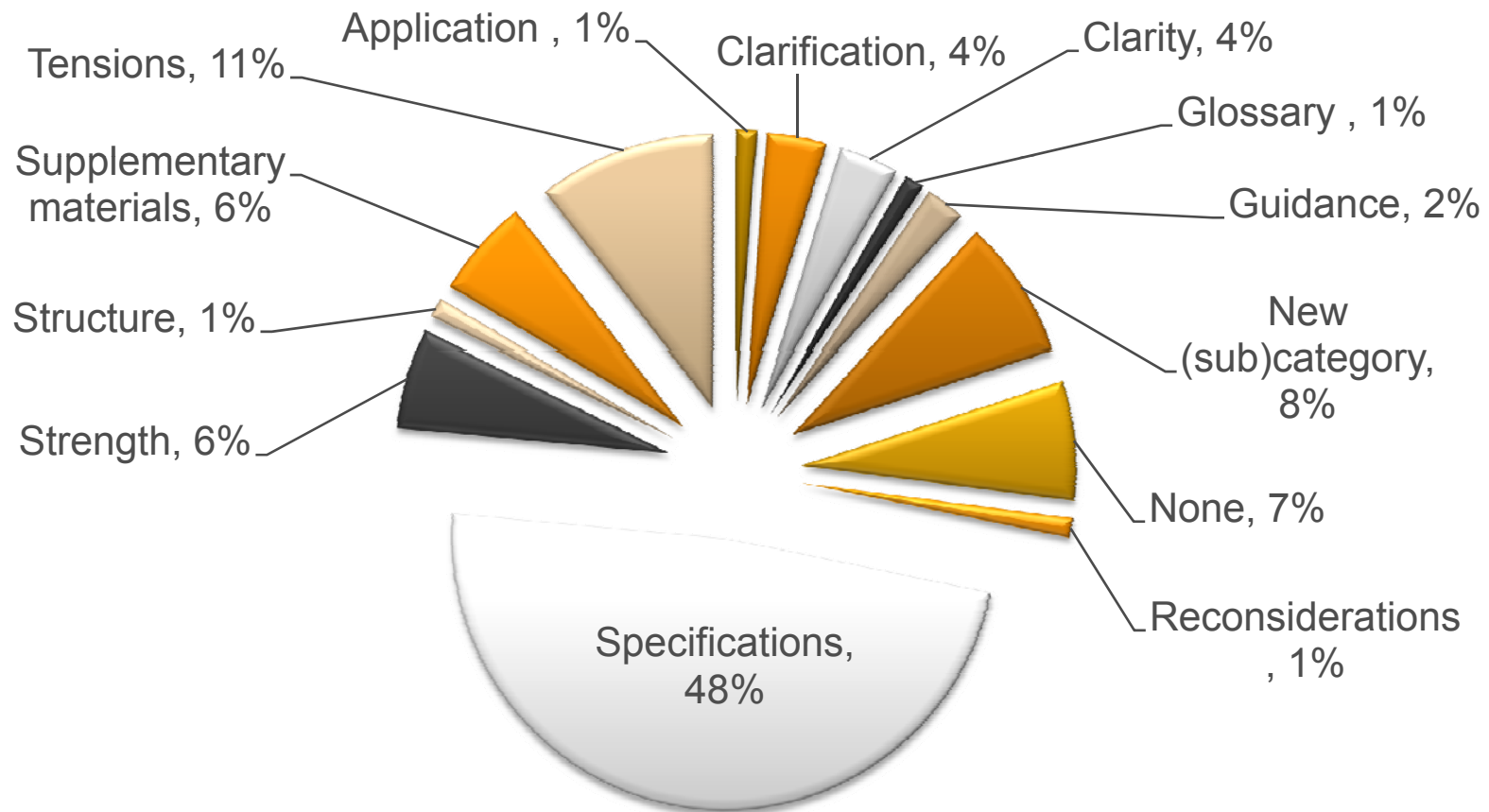


Cross Case Analysis

Item	Question	Responses	Codes across responses	Unique themes
Omissions	What is missing?	85	85	13
Commissions	What components or checkpoints are not necessary?	58	52	9
Confusions	What areas are confusing?	71	68	8
Errors/ problems	Are there any other errors or problems that need to be addressed?	55	50	11
Strengths	What, if any, are the strengths of the checklist?	79	105	12
Suggestions	How can the checklist be improved?	75	94	10

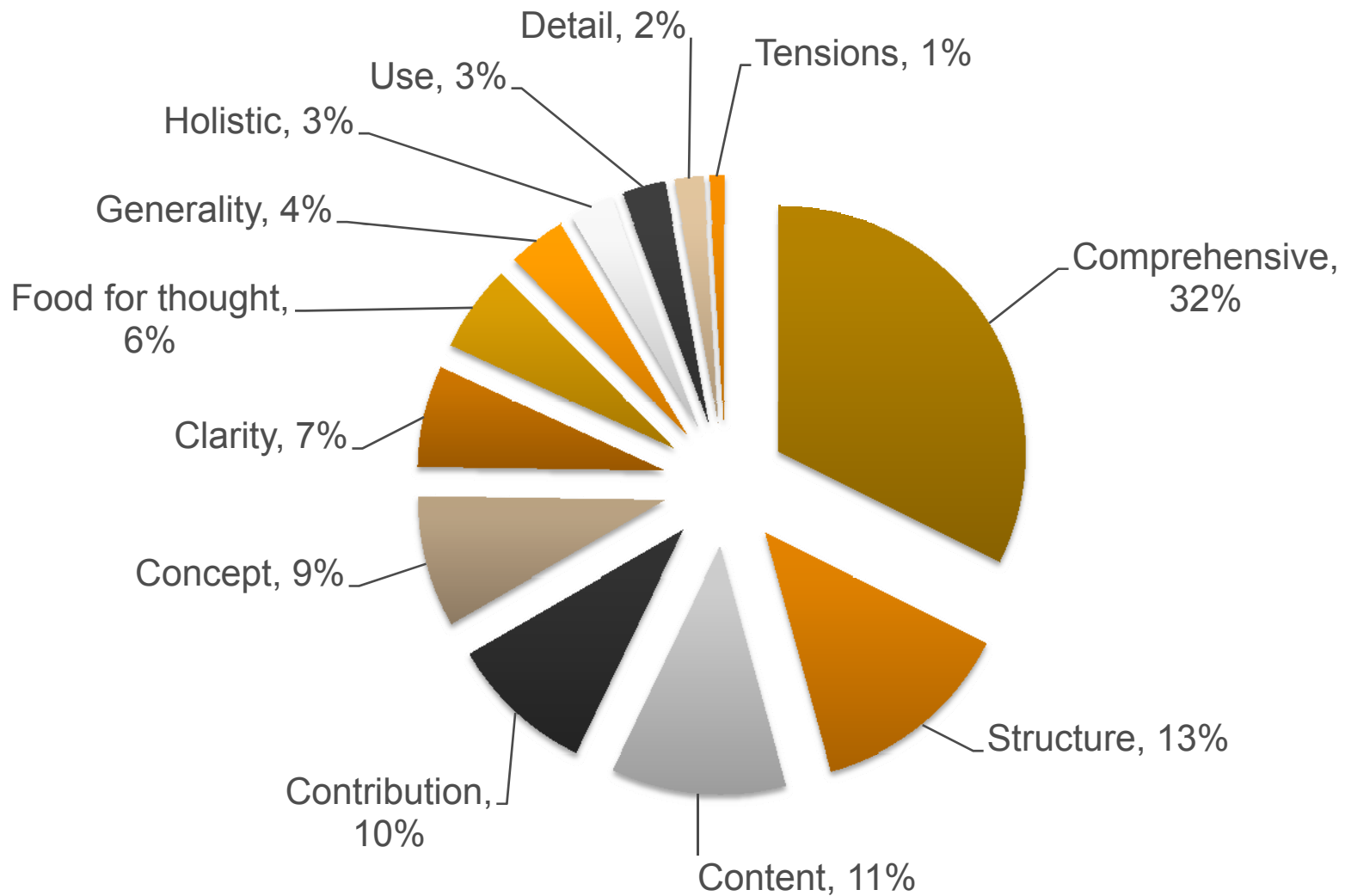


Example “Omissions”

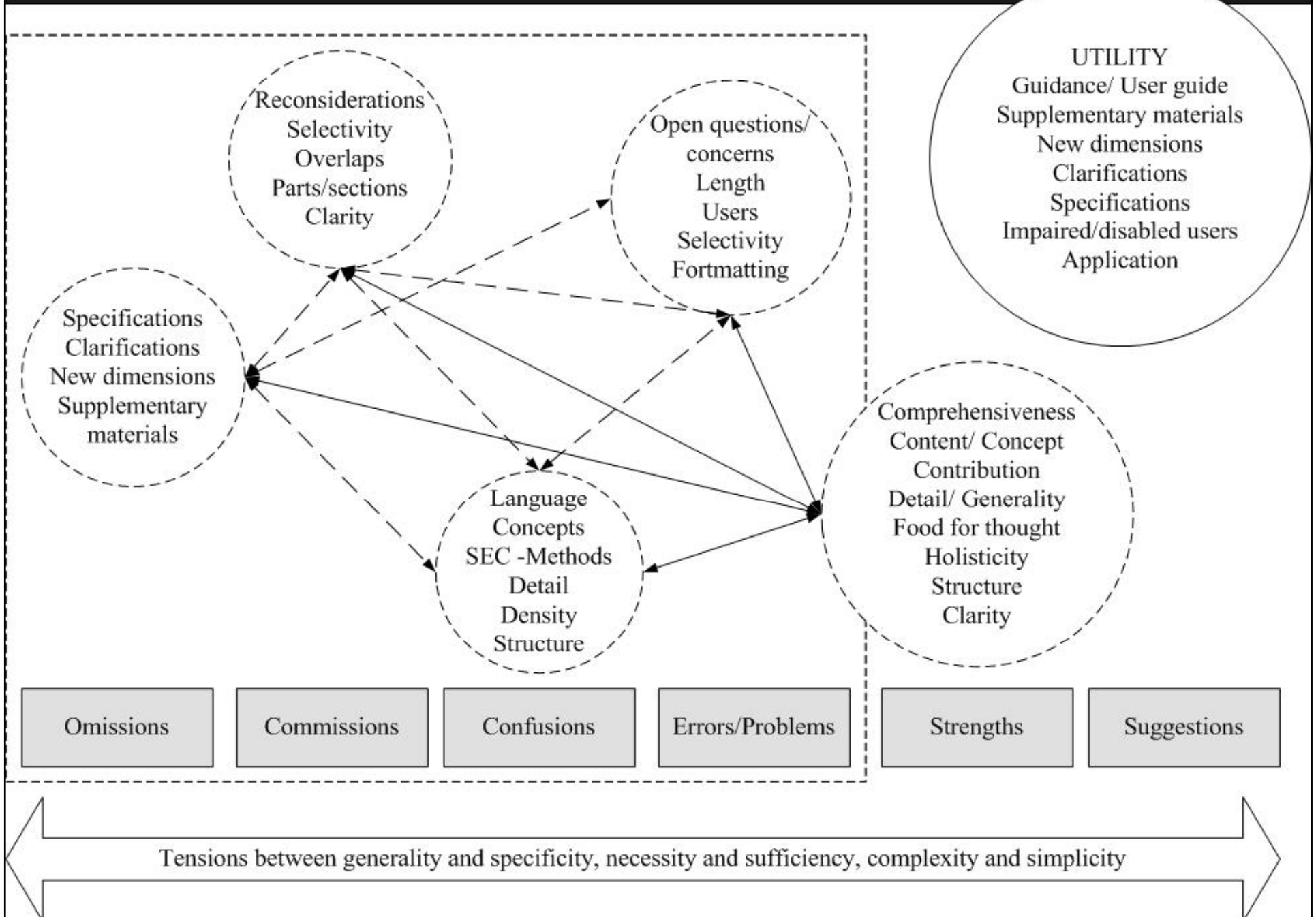




Example “Strengths”



Cross-item Analysis





Cross-section Analysis & Triangulation

- Cross-section analysis
 - Contrasting results by sections (strength, weaknesses, and recommendations)
- Triangulation
 - Comparison of results from the quantitative scales and the qualitative analysis
 - Conclusions for revising the SEC



SEC Validation Method

Strengths & Weaknesses

Strengths

- Relatively quick validation process, but long development process
- Validation based on relevant evaluative criteria and open to critical feedback
- Feedback from a highly diverse group of experts and practitioners

Weaknesses

- Development limited by time and resource constraints
- Validation limited to content and perceived utility, though evidence of use exists
- SEC application limited to volunteers, resulting in lack of systematic information about application



Development and Validation

LESSONS LEARNED



Lessons Learned

- Checklist development should address unique attributes of the evaluand
- Sampling frame is critical
- Checklist validation should be grounded in theory, practice, and use
- Mixed method approach provides increased confidence in validation conclusions
- All checklists are a “work-in-process”



Thank You!

wes.martz@gmail.com

daniela.schroeter@gmail.com