

Fidelity of Test Development Process within a National Science Grant

by

Teresa Brumfield

October 10, 2008



THE UNIVERSITY of NORTH CAROLINA
GREENSBORO

Overview

- Introduction
- Methodology
- Results
- Conclusions
- Recommendations



Introduction

- Statement of the problem
 - Tests: major source of information
 - Problems:
 - Insufficient time and resources
 - Inadequate expertise
 - Conflicting priorities

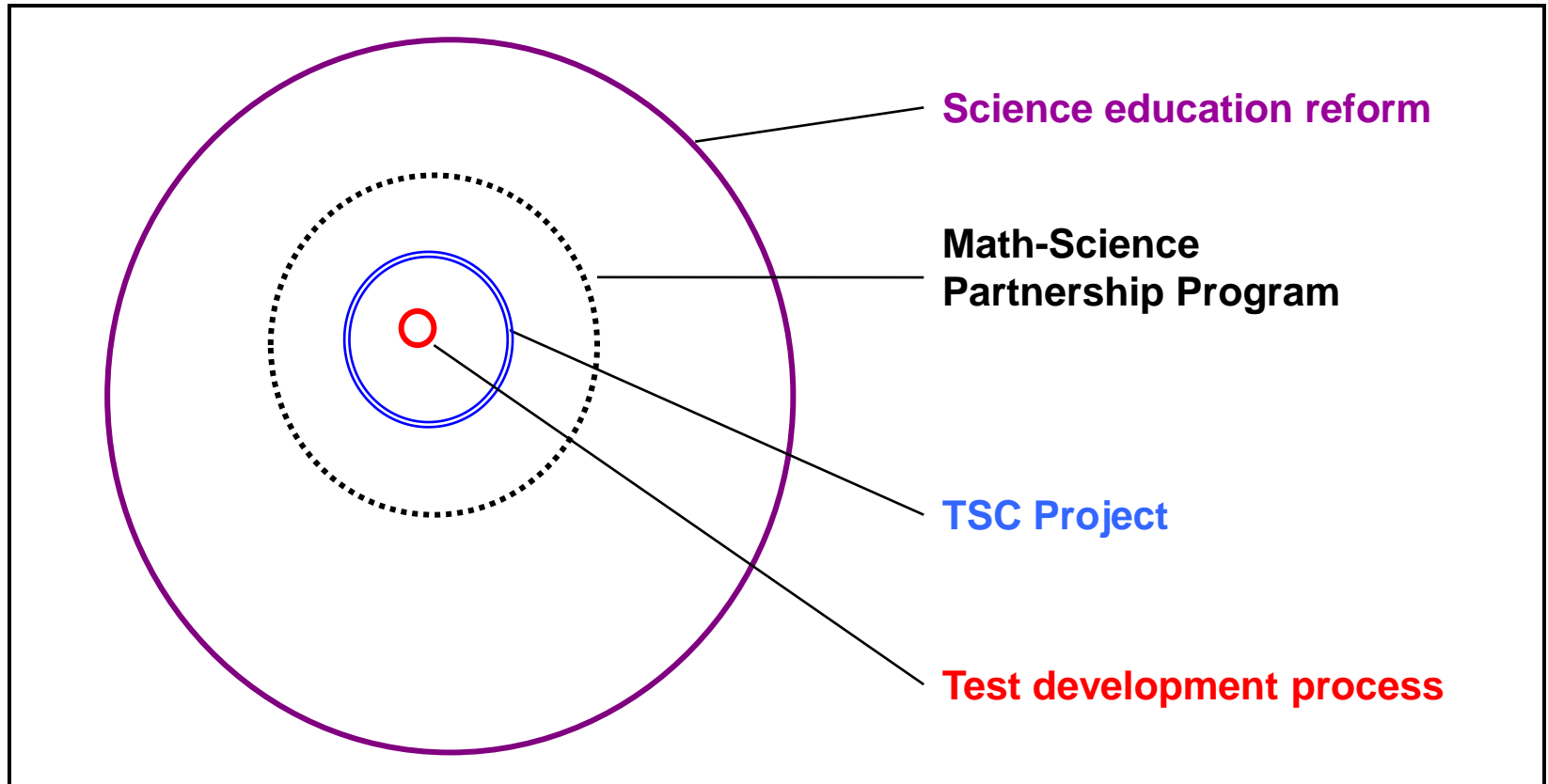


Introduction (*cont.*)

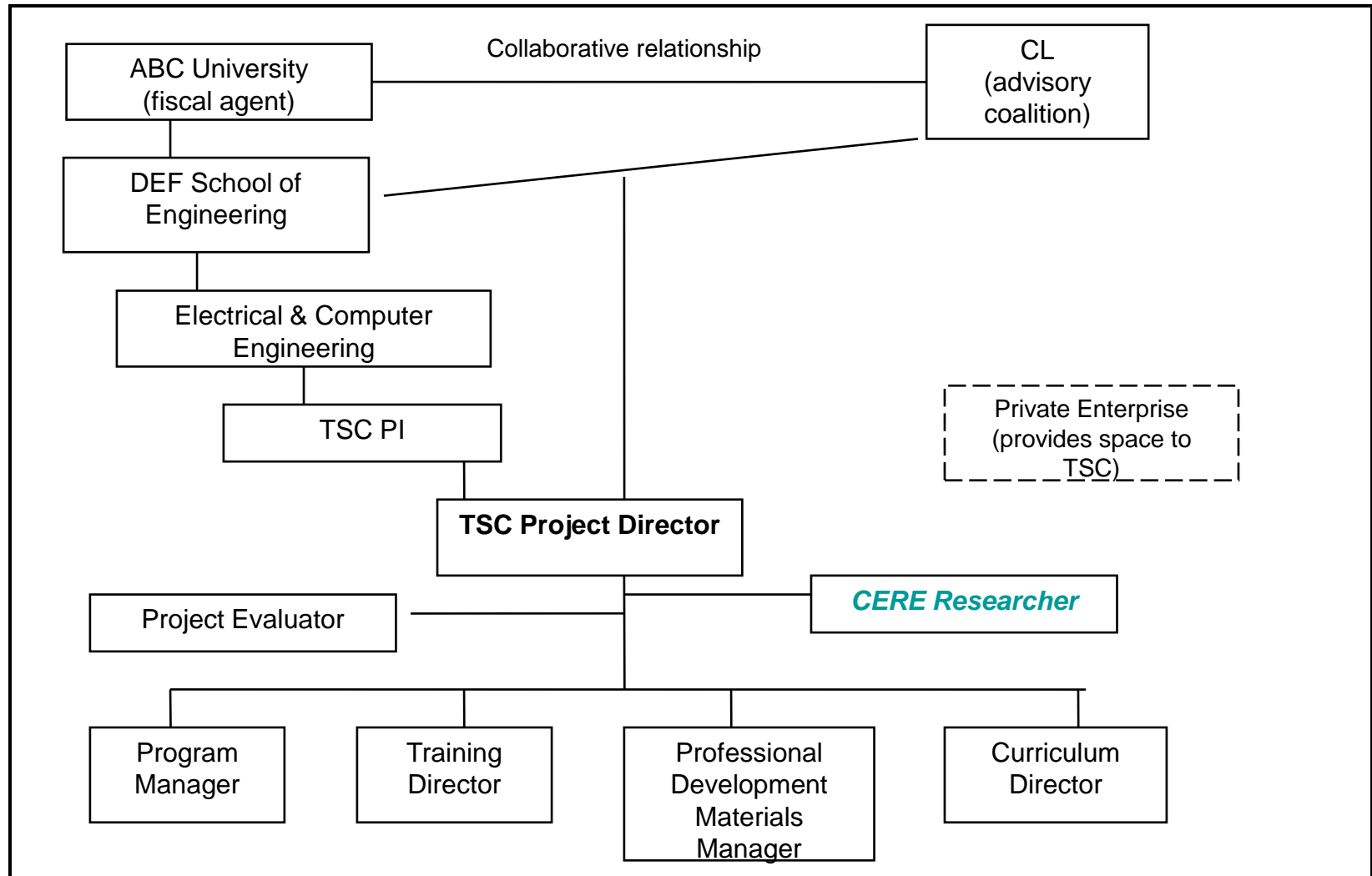
- Objectives
 - Planned vs. actual test development
 - Suggested recommendations



Introduction (*cont.*)



Introduction (*cont.*)



Methodology

- Case study
 - Why case study
 - Unit of analysis
- Data sources and acquisition
 - NSES and 2004 NC Standard Course of Study
 - NSF Program Solicitation 02-061
 - TSC's documents
 - Researcher's notes, personal correspondence, documents, interviews
- Method of analysis



Results

- Planned test development
 - Part 1: contractual agreement between parties
 - Part 2: testing standards (i.e.,
AERA/APA/NCME 1999 Standards for Educational and Psychological Testing)
- Actual test development



Test Development Process

- Planning phases
 - Phase 1 (Test framework)
 - Phase 2 (Test specifications)
- Implementing phases
 - Phase 3 (Pilot test)
 - Phase 4 (Operational test)



Phase 1: Test framework

- Establish test's framework
 - Purpose of the test
 - Scope of the construct (i.e., what it is to measure).
 - Applicable testing standards
 - 3.2
 - 3.11



Actual Test Purpose

- To assess instructional effectiveness of MSP project's professional development
- To demonstrate impact of TSC project



Factors that Affected Phase 1

- National and state science standards
 - Impact on MSP program goals
 - Impact on content of instruction
- NSF's definition of "evidence"
 - *Prefer* quantitative data
- TSC project proposal to NSF
 - Two sets of tests!



Phase 2: Test Specifications

- Develop and evaluate the test specifications, that is:
 - Item formats
 - Response formats
 - Number of items
 - Instructional objectives to be measured
 - Cognitive skills required
 - Time restrictions, if applicable
 - Scoring procedures
 - Test administration procedures
 - Applicable testing standards
 - 3.2
 - 3.6



Actual Test Specifications

- Two-dimensional grid
 - NC Thinking Skill
 - NC SCS science competency goal/objectives
- Estimated percentage of time TSC spent on each instructional objective
- Minimum number of items to be created per instructional objective
- 8 teacher and 8 student MCQs per unit



Test Blueprint (partial)

| <i>Instructional objectives for each grade-level science competency goal</i> | <i>% of time spent on each objective</i> | <i>Minimum number of items to be created for each objective</i> | <i>NC Thinking Skill to be used by test-taker</i> |
|---|--|---|---|
| 1.03 Investigate and describe how plants pass through distinct stages in their life cycle including. <ul style="list-style-type: none">• Growth.• Survival.• Reproduction | 50% | 4 | <ul style="list-style-type: none">•Organizing•Applying•Analyzing•Generating•Integrating•Evaluating |



Factors that Affected Phase 2

- National and state science standards
 - Item type
 - Difficulty level
- TSC project personnel's understanding of:
 - the role the tests were to play in its project evaluation
 - the test development process



Phase 3: Pilot Test

- Construct and evaluate the initial (or pilot) test, that is:
 - Generate items
 - Select items based on criteria
 - Assemble/administer pilot tests
 - Evaluate pilot tests:
 - Item analysis
 - Scoring procedures
 - Test administration procedures
- Applicable testing standards
 - 3.4
 - 3.7
 - 3.8
 - 3.9
 - 3.19
 - 3.22



Actual Pilot Test Development

- Item development
 - Teacher recruitment
 - Item writing workshop
 - Item generation and revision
- Pilot test construction
- Test administration procedures
- Scoring procedures
- Pilot testing, item analyses, test revisions



Factors that Affected Phase 3

- TSC scientists and participating teachers
 - Teacher-item writer recruitment
 - Assumptions by TSC
 - Item generation and revision
 - Assumptions by TSC
 - Teachers' inadequate content knowledge
 - Teachers' inadequate item writing skills



Factors that Affected Phase 3

- Pilot test assembly
 - Compressed time table
 - One version for *both* teachers and students
- Pilot test administration
 - Teachers' and scientists' reluctance
 - Poor data quality
 - Resulting changes
- Pilot test revision
 - Primarily done with TSC project director
 - Relatively problem-free



Phase 4: Operational Test

- Assemble and evaluate test for operational use:
 - Based on pilot test results
 - Revise/replace/delete items
 - Revise scoring procedures
 - Revise test administration procedures
 - Assemble, administer , and evaluate operational test to intended population of test-takers
 - Applicable testing standards
 - 3.19
 - 3.20



Actual Operational Test

- No provision in TSC-CERE subcontract



Conclusions

Constructing psychometrically sound tests within an evaluation is *challenging*:

- Sufficient time and resources seldom provided (see also, Rossi, et al., 2004)
- Not routine or unproblematic (see also, Wolf & Cumming, 2000).



Recommendations

- Familiarity with test development steps and standards
 - MSP project directors
 - Project evaluators
 - NSF program officers
- Identification of *all* stakeholders
- Future research
 - Meta-analysis
 - Alignment protocol

