Stage 1: Develop Scoring Guides

- Scoring guides developed along with items
- Refined during item reviews:
  - Item development contractor
  - Standing committee (content experts)
  - State item review
  - Ongoing NCES reviews

Stage 2: Score Pilot

- Pilot scoring guides and training packets reviewed by NCES and standing committee
- Submit data for item analysis
- Select items for operational assessment
- Refined during item reviews:
  - Standing committee
  - Ongoing NCES reviews

Stage 3: Score First Operational Administration (or Pre-Calibration)

- Operational scoring guides and training packets reviewed by NCES and standing committee
- Submit data for analysis

Stage 4: Score Operational Administration (Trend Scoring)

- Submit data for analysis

Multiple-Choice Items

- Scanned and processed electronically with quality control and validity checks

- All Other Subjects

- Scaling and reporting

Reading and Mathematics

- Item pre-calibration

- First operational administration

- First Operational Administration (or Pre-Calibration)

- All Other Subjects

- First Operational Administration

- All Other Subjects

- Scaling and reporting

- Second operational administration

- First Operational Administration (or Pre-Calibration)

Phase I.
Scoring Guide Development and Pilot Scoring

- NAGB Reviews

Phase II.
First Operational Scoring (or Pre-Calibration)

- NAGB Reviews

Phase III.
Subsequent Operational Scoring

- NAGB Reviews

- Constructed-Response Items

- Scanned responses scored by qualified and trained scorers using an electronic image-processing and scoring system

- Content experts check student papers against scoring guides

- Refine scoring guides

- Select examples and practice papers for training packets

- Train scorers

- Ensure scorer quality

- Monitor scoring accuracy and reliability
  - Individual scorers (backreading)
  - Between scorers (inter-rater reliability)
    - (25% of responses double scored)

- Conduct Scoring Debriefing and Document Potential Refinements to Items and Scoring Guides

- Content experts check student papers against scoring guides

- Refine scoring guides

- Select examples and practice papers for training packets

- Train scorers

- Ensure scorer quality

- Monitor scoring accuracy and reliability
  - Individual scorers (backreading)
  - Between scorers (inter-rater reliability)
    - (5–25% of responses double scored)

- Archive Final Scoring Guides and Training Materials

- Content experts check student papers against scoring guides

- Refine scoring guides

- Select examples and practice papers for training packets

- Train scorers

- Ensure scorer quality

- Calibrate scoring

- Monitor scoring across-year consistency

- Meet criteria for trend

- Update Documentation on Scoring and Training After Each Administration

April 2005