The Duplex Design in the 21st Century

Isaac I. Bejar

Presented at
National Conference on Student Assessment

June 21, Orlando, Florida.

Unpublished Work Copyright © 2010 by Educational Testing Service. All Rights Reserved. These materials are an unpublished, proprietary work of ETS. Any limited distribution shall not constitute publication. This work may not be reproduced or distributed to third parties without ETS's prior written consent. Submit all requests through www.ets.org/legal/index.html.

Educational Testing Service, ETS, the ETS logo, and Listening. Learning. Leading. are registered trademarks of Educational Testing Service (ETS).
Overview

- General test-based accountability requirements
- Race to the Top requirements
- The duplex design
- Augmented by recent advances
- Conclusions
High-level design


Details change with each new cycle, e.g., Race to the top

• Measures student knowledge and skills against a common set of college- and career-ready standards in mathematics and English language arts in a way that
  1. Covers the **full range of those standards**, including standards against which student achievement has traditionally been difficult to measure;
  2. As appropriate, elicits **complex student demonstrations** or applications of knowledge and skills;
  3. Provides an **accurate measure** of student achievement across the **full performance continuum**, including for high- and low achieving students; and
  4. Provides an **accurate measure of student growth** over a full academic year or course

Additional Race to the Top details

5. Are administered at least once during the academic year in grades 3 through 8 and at least once in high school

6. Produce student achievement data and student growth data that can be used to determine whether individual students are college- and career-ready or on track to being college- and career-ready

7. Produces data, including student achievement data and student growth data, that can be used to inform determinations of school effectiveness for purposes of accountability under Title I of the ESEA
Translation into design options

<table>
<thead>
<tr>
<th>Race to the Top requirement</th>
<th>Design options</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consecutive grades 3-8; On track; Growth within year</td>
<td>Coherence of performance standards across grades, and occasions within grade -&gt; different approach to standard setting</td>
</tr>
<tr>
<td>Complex responses</td>
<td>Balance CR and shorter items, scoring and timeliness potential issue -&gt; automated scoring</td>
</tr>
<tr>
<td>Accuracy across performance continuum</td>
<td>Test too long -&gt; a common adaptive testing,</td>
</tr>
<tr>
<td>Multiple states</td>
<td>Security, capacity, multiple forms -&gt; comparability within year and year-to-year</td>
</tr>
<tr>
<td>Determinations of school effectiveness; Full range of those standards</td>
<td>High accuracy of student-level measurement -&gt; matrix sampling for comprehensive coverage of standards</td>
</tr>
</tbody>
</table>
Some conditions for comparability
(Harvested from NCME 2011 Marion, Kolen, Lane, Martineau)

- Specified test administration windows, procedures, accommodations, inclusion, security protocols
- Specific computer hardware and software (for CBT and CAT), or at least a narrow range of specifications.
- Common test blueprint
- All states have theoretically promised to adopt common performance-level descriptors and common achievement cutscores.
The duplex design

- Two-stage adaptive test and matrix sampling
- Assumes IRT
- Precedes standards-based reporting
- Precedes automated scoring and item generation


For each grade:

- First and second stage optimized to classify students into achievement levels
- Third stage has multiple uses
  - Growth
  - Complex responses
  - Additional standards
• Not limited to multiple-choice IF automated scoring is feasible
• Many issues revolving around security and capacity
1. Form models (blueprints), populated by item models
2. Automated scoring
Recent experience

• NCARB: Licensing of architects requiring complex simulations
• New GRE: Uses variants of writing prompts
• Constructing parallel simulation exercises for assessment centers and other forms of behavioral assessment (Brummel, Rupp, Spain, Personnel Psychology, 2009)
• Comparability across years and within year
• Coherence across grades and occasions
Conclusions

• The requirements of Race to the Top assessments call for different designs
• The duplex design is an assessment architecture that could accommodate many of these requirements, especially now under computer-based testing
• Recent developments, in item generation and automated scoring, could be make the duplex design more useful