Validity, Test Use and Consequences:

Reflections on Some Current Issues in Educational Assessment

Madhabi Chatterji

TEACHERS COLLEGE
COLUMBIA UNIVERSITY

Validity, Test Use and Consequences:
Reflections on Some Current Issues in Educational Assessment

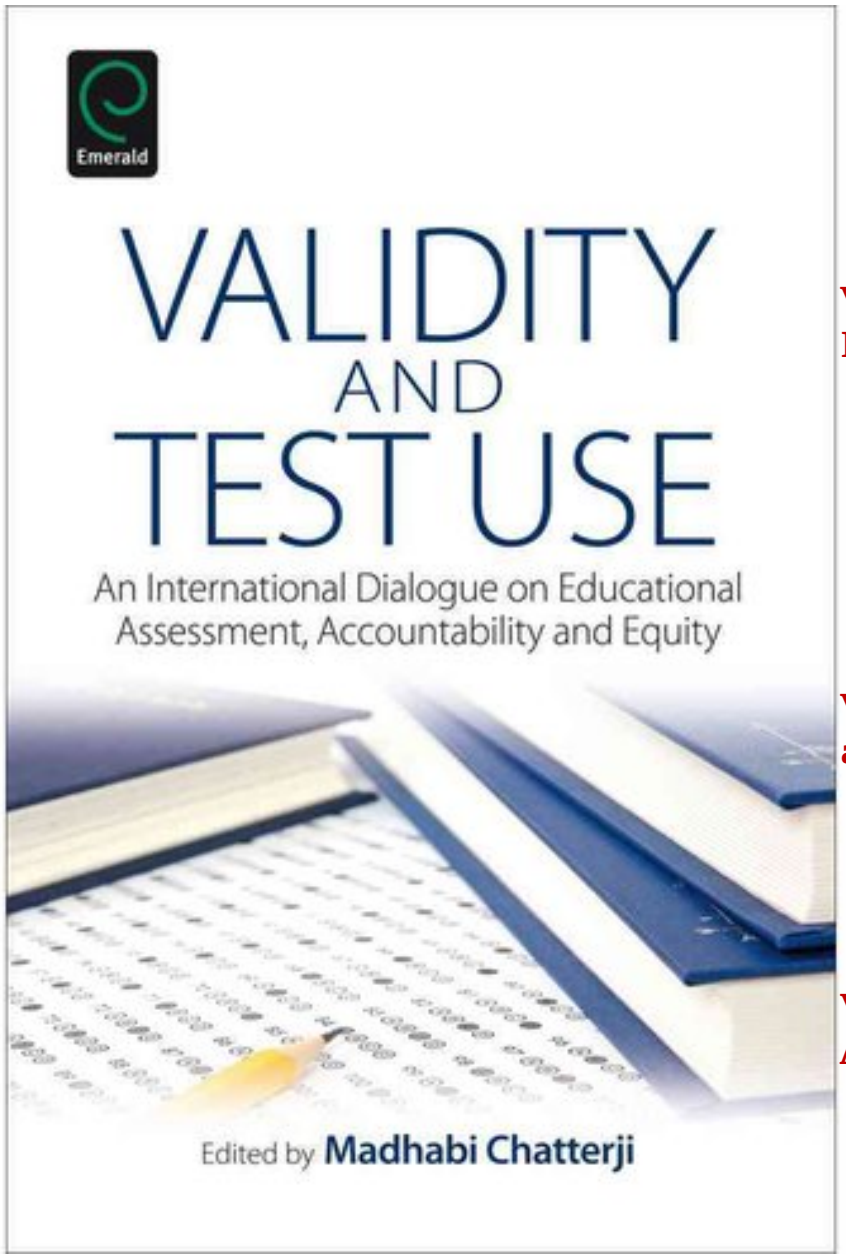
Why a Focus on Validity?

Exponential rise in K-12 testing and accountability programs globally—in both developed and developing regions

Just released PISA 2012 Results for the U.S. and Public Concerns…

Common Core Standards-based Testing and Backlash to the Testing….

- ***What do the test results really mean?***
- ***Are the common interpretations and uses of test-based information meaningful and defensible (i.e., valid) ?***
- ***What can we do to optimize validity in "test use" settings prevalent around the world, and promote more fair and equitable assessment practices?***
- ***How can testing and assessment better serve education (upholding principles of validity)?***



I will provide a quick preview of just-published findings and recommendations from AERI's inaugural conference in 2012:

Validity, Fairness and Testing of Individuals in High Stakes Decision–making Contexts

Lead: **Michael T. Kane**

Reactions: **Alina von Davier, Kadriye Ercikan and Maria Elena Oliveri, Sebastien Georges, Edmund W. Gordon**

Validity Issues in Models of Teacher Evaluation and School Accountability

Lead: **Adrie Visscher**

Reactions: **Jakob Wandall, Aaron Pallas, Drew H. Gitomer, Haniza Yon**

Validity Issues in International Large-Scale Assessment (ILSA) Programs

Lead: **Michael J. Feuer**

Reactions: **Hans Wagemaker, Eduardo Backhoff, Valena White Plisko, Robert Laurie**

Foreword and Concluding Chapter:

When Education Measures Go Public – Stakeholder Perspectives on How and Why Validity Breaks Down
Edited by Madhabi Chatterji
Volume 115, Number 9, 2013

Teachers College Record
The Voice of Scholarship in Education

Companion volume that examines validity issues in educational testing and test-based educational accountability programs in the U.S.

Lead: **Eva L. Baker**
Reactions: **Leo Casey, Jeffrey Henig, David Steiner, Kevin Welner**
Added Commentary: **Lorrie Shepard**
Foreword: **Madhabi Chatterji**

AERI-NEPC Policy-briefs for practitioner audiences ….forthcoming, January, 2014 in *Quality Assurance in Education*.

Premise of the collected works:

Need for a systematic look at how well validity holds up (or breaks down) in practice, policy, and evaluative settings around the world, and conditions and factors that compromise validity, posing recurring validity challenges-- mostly still undocumented or unclear

Validity, Test Use and Consequences:
Reflections on Some Current Issues in Educational Assessment

What makes an achievement test or test-based information valid?

- **Built-in features of an achievement test—such as:**
  - **Items that are well written to match the curriculum**
  - **Scoring systems that are reliable**
  - **Metric of measurement that is meaningful for users.**
- **Also depends on how appropriately the test and test-based information are used and interpreted in applied settings**
  - **The extent to which *actions* or decisions taken with test scores and test-based reports are supported with evidence of validity.**

(after AERA, APA, & NCME, 1999; Messick, 1989; Kane, 2006; Kane , 2013)

Validity, Test Use and Consequences: Reflections on Some Current Issues in Educational Assessment

My Role:

1.Frame the problem with case studies of validity issues, "breakdowns"

2.Content analysis of main ideas offered in chapters : What can be learned about how validity challenges arise in the test use contexts examined? What are the suggested recommendations, frameworks, solutions of authors to improve the *status quo*?

- **Findings of content analysis were organized by theme**
  - **Type of validity challenge or oversight?**
  - **Specific example in test use setting**
  - **Who could face consequences?**
  - **What could test developers and researchers do to improve validity?**
  - **What could test users do to improve validity?**
- **Which themes are triangulate and generalize across test use contexts?**
- **Which themes are unique to particular test use settings?**
- **My own recommendations**

Validity, Test Use and Consequences:
Reflections on Some Current Issues in Educational Assessment

Starting Definitions, Premises and Parameters:

Validity-

meaningfulness of scores and test-based reports and defensibility of interpretations and actions taken in test use contexts

Validation-

formal processes for collecting evidence to support inferences and actions to be taken with scores and test-based reports

Stakeholders- test developers, psychometricians and affiliated researchers, test users—educators, policy-makers, parents, students and public

If we accept the above definitions:

Validity is a shared responsibility!

Findings and Recommendations-One Issue evidenced in School and Teacher Evaluation Contexts

Example of an Identified Validity Challenge or Threat:

Same test(s) used to serve multiple purposes

- **Conflicting assessment purposes, "theories of action" (rationale for testing), values, and information needs of different assessment stakeholders at different levels of *same* education system**

Consequences:

- **Strong design and validity evidence to support some score interpretations and test-based information uses (e.g., student learning at classroom level);**
- **Weak or no validity evidence for others (e.g., school or teacher quality at upper organizational level)**
- **Undesirable practices/policy actions--unintended consequences in high stakes settings, some unforeseen**

Does the issue generalize internationally?

Found in the U.S. and to some extent, internationally

Specific Case Illustrations:

The Netherlands -discussed in the *Validity and Test Use* book's Section 2

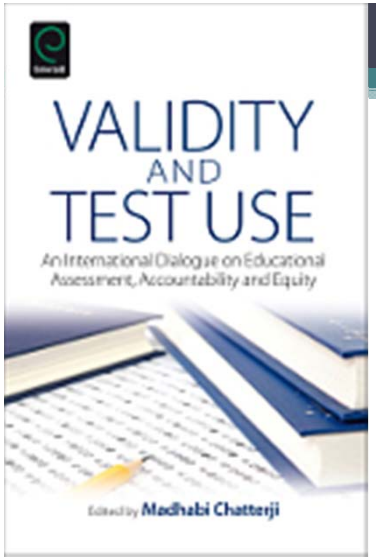
U.S. Common Core Initiative- my extrapolation of what we learned to new case

The Netherlands' Evaluation-centered Primary School Model

Achievement Oriented Work (AOW) Policy

- **Formative assessment with standardized achievement tests to guide classroom instruction and student learning-- as a part of everyday work culture**
- **Results aggregated up and made public for accountability purposes**
- **Relatively low stakes tied to test results for schools and teachers**

Visscher, 2013; Gitomer, 2013; Pallas,2013; Wandall, 2013; Yon, 2013



Logic Model for the Netherlands' Achievement Oriented Work (AOW) Policy in Primary Schools

Purpose for testing at the classroom level:

Formative assessment to improve instruction and student learning (Visscher, 2013)

Validity of test-based information for inferences on student achievement:

Relatively High

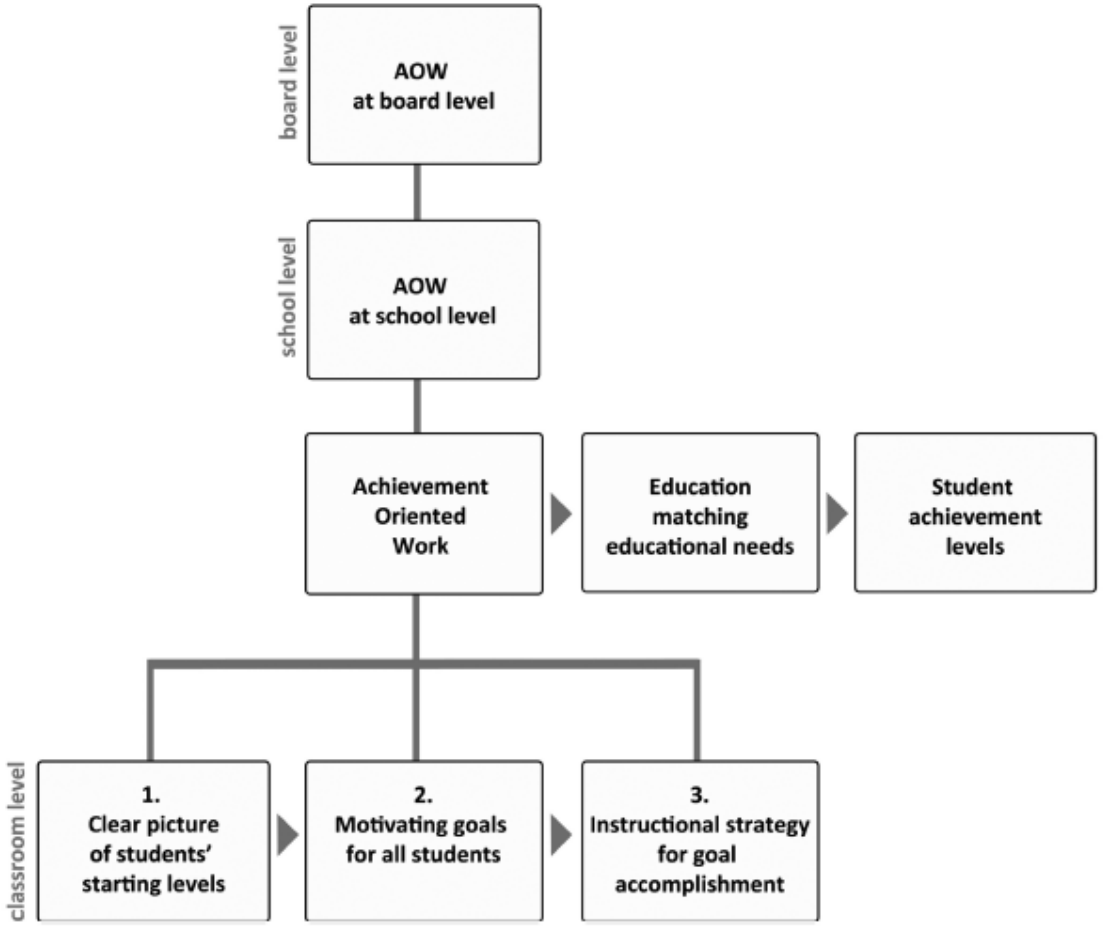
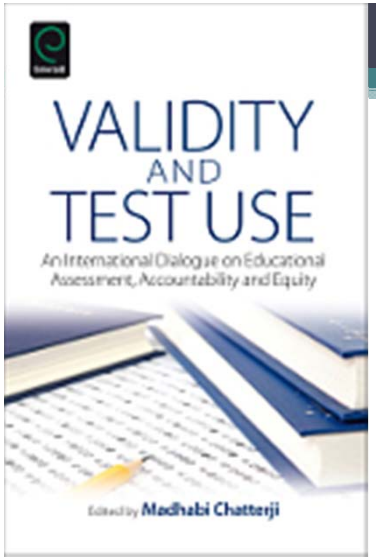


Figure 7.2: Components and logic model showing anticipated effects of AOW policy in Dutch primary schools. *Source*: Adapted from Visscher & Ehren (2011).



Logic Model for the Netherlands' Achievement Oriented Work (AOW) Policy in Primary Schools

Purpose for testing at upper organizational level: Public Accountability

Validity of test-based information for inferences on school quality: LOW

BUT, current accountability requirements are mild and "low stakes" in the Netherlands. Hence, less consequential for schools and teachers. (Visscher, 2013)

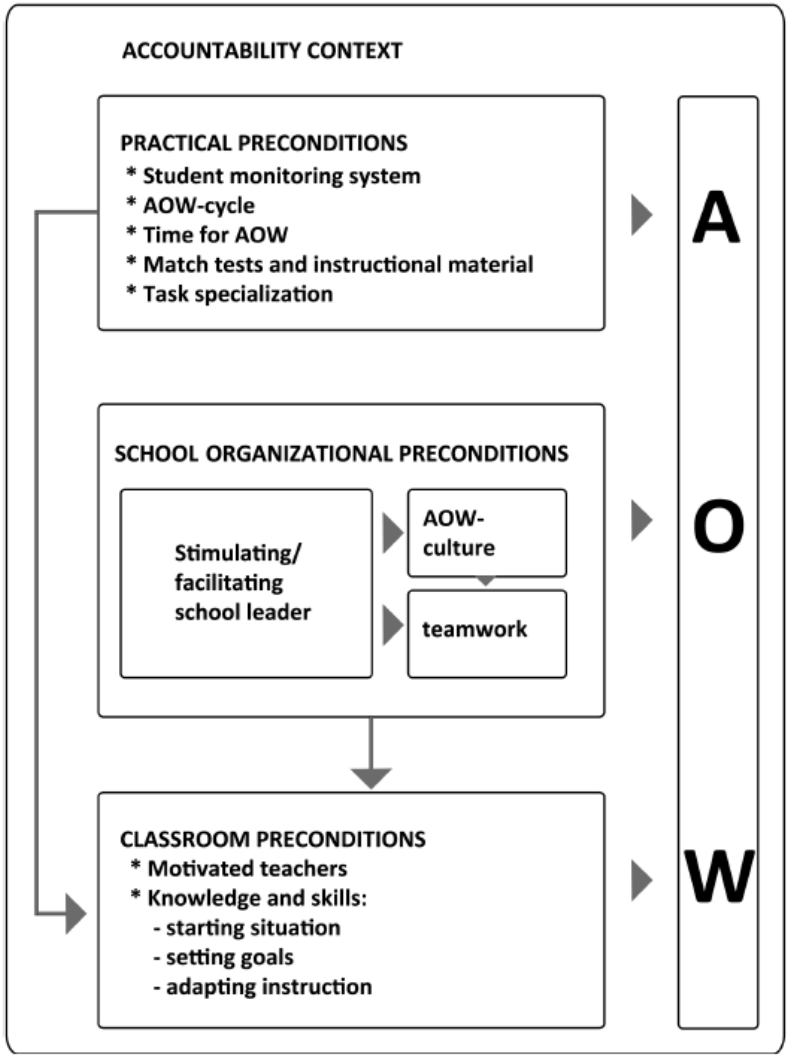


Figure 7.3: Prerequisites for implementing AOW policy in schools.

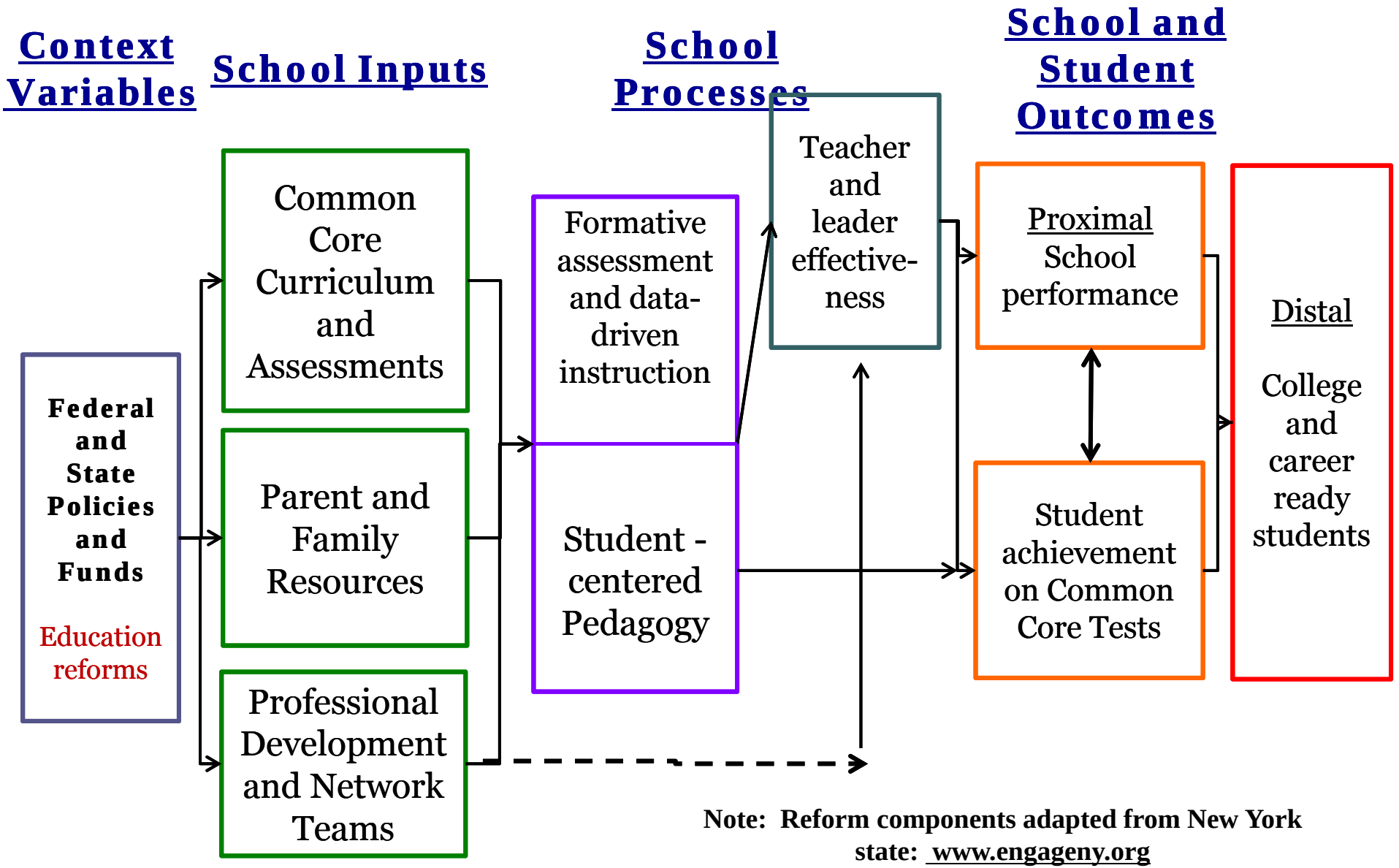
U.S. Case- Common Core Initiative

- **Well-intentioned goals and mission**
- **Curriculum and tests designed to match higher standards, implemented with a long-term view to prepare students for college and career success**

BUT:

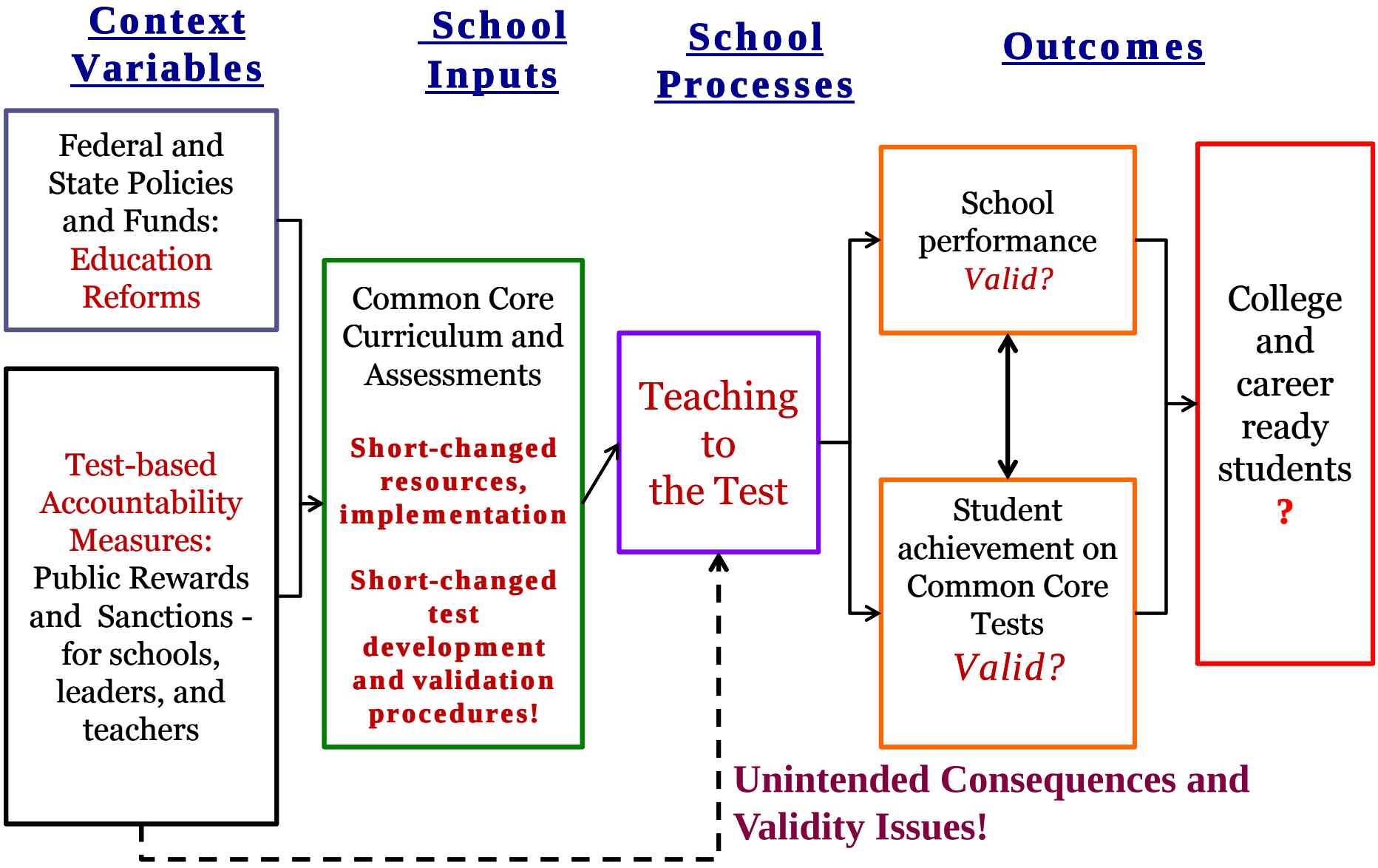
- **Tests used as a top-down policy tool to implement reforms before reform components are in place (Welner, 2013)**
- **High stakes tied to results of student testing for schools, teachers, leaders**
- **AND:** ***Unrealistic timeline for policy implementation, with validity threatened***

Common Core Initiative -In Theory



Note: Reform components adapted from New York state: www.engageny.org

Common Core Initiative -In Practice



Validity, Test Use and Consequences: Conclusions

Validity, Test Use and Consequences are Inseparable

What We Need: More stakeholder dialogue and cross-learning from the earliest stages of test development, validation and use to address and preempt issues documented

What could test developers and assessment researchers do?

- **Apply broader validation frameworks that engage stakeholders and policy-makers with testing experts to anticipate and examine issues (Kane, 2013; Feuer, 2013).**
- **Make reports and validity evidence understandable to public and lay users.**
- **Use systems-oriented logic models to create coherent frameworks to guide design of assessment programs, nested within larger education systems**

What could test users do?

- **Become more informed consumers of tests and test-based information**
- **Make decisions supported with validity evidence (Baker, 2013; Shepard, 2013)**
- **Curb over-interpretation or over-ambitious use of test or test-based information**