

# Theories of action as a tool for explicating policy goals and intended uses of assessments



**SCOTT MARION**  
**CENTER FOR ASSESSMENT**  
**REIDY INTERACTIVE LECTURE SERIES**  
**CAMBRIDGE, MA**  
**OCTOBER 21-22, 2010**  
**[WWW.NCIEA.ORG](http://WWW.NCIEA.ORG)**

# Purposes & Uses

2

- **Axiomatic to say that purposes and uses are critical in evaluating the validity of assessment and accountability system**
- **Yet, the current policy context is requiring (or promising) assessments to serve multiple purposes**
  - e.g., to support teacher evaluation systems, provide formative information, evaluate whether students are “on track towards college/career readiness.”



# Focus of talk on “non-tested”

3

- Tension between using student assessment data for teacher evaluation purposes while hoping to extract instructional information from these same data
- States are wrestling with how to incorporate student achievement and growth data from “non-tested subjects and grades” in teacher evaluation systems
- I will be over-simplifying the example to help make some points to demonstrate how a theory of action can help clarify policies and policy alternatives



# Theory of Action

4

- A theory of action is a useful starting point for developing a validity argument and validity evaluation plan
- I use the “non-tested” subjects and grades issue as an example to demonstrate how the careful explication of a theory of action can clarify the intended uses and make clear the mechanisms required for such uses to be fulfilled
- The theory of action should also help reveal potential irresolvable conflicts in proposed uses.



# Thinking Through a Theory of Action

5

- Policy makers should have to very explicitly say why implementing test-based approaches to support educator effectiveness for these grades and subjects will lead to improved educational opportunities for students
- In addition to the why, policy makers should have to describe the **how, or the mechanisms** by which they think that these improved learning opportunities will occur
  - For example, one might postulate that holding teachers accountable for increases in student test scores on classroom-based assessments will lead to the development of both better assessments and improvements in student learning.
- The evaluator and/or user must specify the mechanism by which these accountability uses will lead to the anticipated changes in teaching practices, such as targeted instruction and/or more appropriate curricular materials.



# We need a comprehensive approach

6

- While I use the “non-tested” subjects and grades as the content of my example, my colleagues and I argue that it will be more fruitful to comprehensively approach the issue of student outcome-based teacher accountability (status, growth, VAM)
- “Tested” and “non-tested” subjects and grades can then be viewed as special cases of the comprehensive framework



# The Challenge of Non-Tested Grades

7

- Advances in growth and value-added models has increased interest in using student test scores as part of educator accountability systems
- Education leaders, especially those submitting Race-to-the-Top (RTTT) applications, have quickly realized that evaluation systems focusing on teachers in subjects and grades for which there are state tests generally means that only one-quarter or so of the teaching force would be subject to such evaluations



# Just Add Tests....

8

- **Problem**: Few, if any, technically adequate tests for these non-tested subjects and grades
- **Solution**: Spend enormous sums of money to provide external tests in science, social studies, arts, PE, etc to provide the data necessary for calculating growth or value-added quantities
- **Reality**: Of course, this isn't really a solution for both pragmatic and technical reasons





## Some “straw men”

9

- States and others have proposed approaches that revolve around using aspects of the classroom-based assessment system
  - Feeding the data from these classroom-based measures into some sort of analytic method used to calculate growth or a value-added quantity
  - Use classroom-based and/or other information to **establish goals** for either individual students and/or the class as a whole
- A desire to continue to use these assessments as part of the teaching and learning cycle



# Theory of Action as a Tool

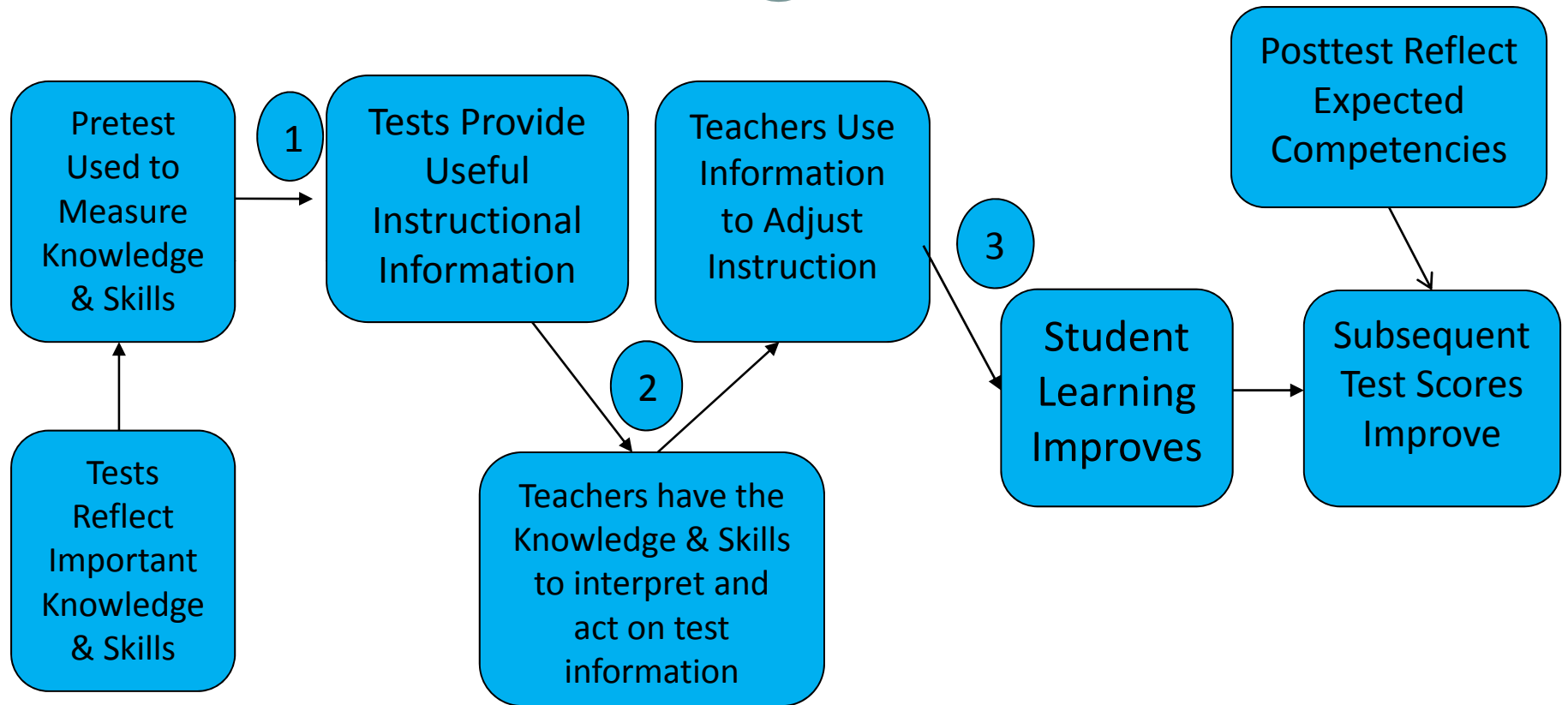
10

- **The following figures attempt to explicate the tension between using assessments for instructional and accountability information**
  - This set of representations assumes at least one pretest and at least one posttest
  - Limited in that I cannot reveal all of the important mechanisms or processes that critically connect among the highlighted components
- **It is the differences in these mechanisms that make the two theories of action irreconcilable**



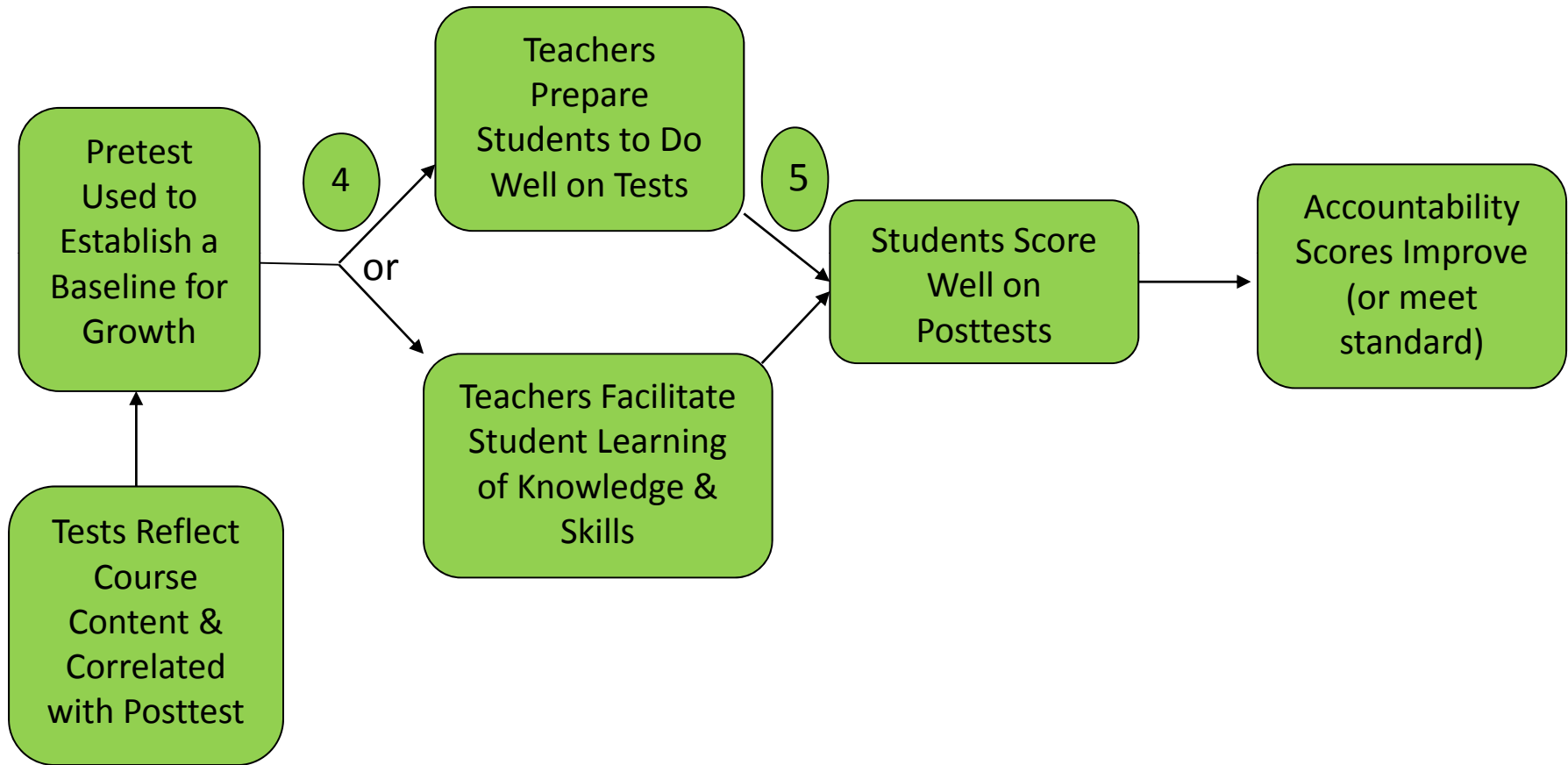
# Theory of Action for Local Tests to Inform Instruction

11



# Theory of Action for Local Tests Used for Accountability

12



# The “pretest”

13

- **Instructional**: important that the pretest measure either the expected prior knowledge or the forthcoming required knowledge and skills
  - provide insights for the teacher as she plans her upcoming instructional activities (see #1)
- **Accountability**: simply requires that the pretest be related to the posttest, but in reality it should be conceptually related to the posttest to make for a more valid accountability determination, especially if it is “growth-based”.



# Increasing knowledge or test scores?

14

- **#4 in the accountability ToA represents an important choice point for many educators**
  - Rarely explicit because many naively believe that both options are equivalent
- **A major concern when assessments intended to support instructional uses are shifted to accountability uses**
  - Teachers may shift their focus as expected, depending on the stakes, to increasing test scores instead of ensuring that students are learning the content at the intended depth



# Some important claims for validation

15

- When tests are designed to support instruction, the test (and reports) must yield information relevant to instruction
  - Teachers must be able to interpret this information **and** know what to do next in terms of instruction (see mechanism #2).
- Both aspects of the ToA require compelling evidence to support claims that assert, for example, that teachers possess pedagogical content knowledge adequate to properly interpret student performance information and make appropriate instructional adjustments.



# Moving toward outcomes

16

- **Instructional ToA: prior actions will lead to improved student learning (mechanism #3)**
  - Should lead to higher test scores
- **Accountability ToA assumes only that students test scores will increase compared to the pretest (mechanism #5).**
  - Increased test scores should lead to higher accountability results which are intended to motivate continued and improved actions by teachers





# Reconciling the theories of action

17

- This close examination makes it appear that the ToA are too divergent to find any reconciliation
- However, given the intense interest among many policy makers to find an “answer”, let’s take another look to search for possible reconciliation
- In this case, the instructional theory of action is considered the “primary use” since that is the current use case, but we will work to see if accountability uses can be accommodated.



# Pretest-Posttest

18

- **On the surface, evaluating the pretest-posttest (or some analytic analog) for the entire course would not disrupt the instructional purposes**
  - Could still allow for accountability determinations
  - In a perfect, non-corruptible world, this might be true, but let's examine some of the measurement issues before getting into the shortcomings of human beings



# Conditional Status $\neq$ Growth

19

- **Instructional**: Should help determine students' knowledge and skills... provide useful information for planning instruction.
  - Correlation with end of year test is NOT a requirement
- For example, a high school chemistry pretest should evaluate students' understanding of proportional reasoning (a critical math skill for high school chemistry) to see what type of early mathematics remediation is necessary
- **Accountability**: Some sort of mini-version of the end of course test might serve as a fair accountability pretest so we may judge how much students have learned relative to the expected domain.
  - Hard to see how having students get a lot of wrong answers because they have not had an opportunity to learn the material would serve instructional purposes.
- Might be less challenging to bridge the two sets of uses in a domain organized by a shared learning progression
- **Conditional status** is not the same as **growth**



# Major Conflict

20

- **Problem:** Many conflate test scores with student learning
- **Solution:** If teachers in both contexts focused on increasing student learning, the two uses could be bridged
- But reality strikes...We must surface the history of accountability testing
- Naïve to think that teachers, knowing they are being held accountable for increasing students' scores on specific tests, would not focus their instruction on what they expect to be covered on the accountability test.



# The Outcome Measures

21

- Little doubt that a carefully designed assessment could possibly serve both uses, but for the sake of argument...
- **Accountability**: Depending on the stakes, the accountability test might place a premium on reliability
- **Instructional**: Designed to provide rich evaluative information and (hopefully) probe the depths of students' understanding. Reliability would be less of a design concern than construct validity.
- This example is purposely exaggerated—this might be a place to try to use the same thoughtfully designed assessment in both cases



# Pre-Post Summary

22

- It might be possible to find efficiencies by using the same assessments to serve both purposes, but it is unlikely that some of the fundamental differences will be satisfied such that a single assessment system can serve both purposes well.
- I even downplayed the likely corruption when teachers are being held accountable for changes in test scores that are administered as part of their regular classroom experience.
  - This actually swamps any of the measurement concerns
- The theory of action exposes many of the components and mechanisms where teachers could be tempted to act in less than honorable ways



# The Goal Setting Example

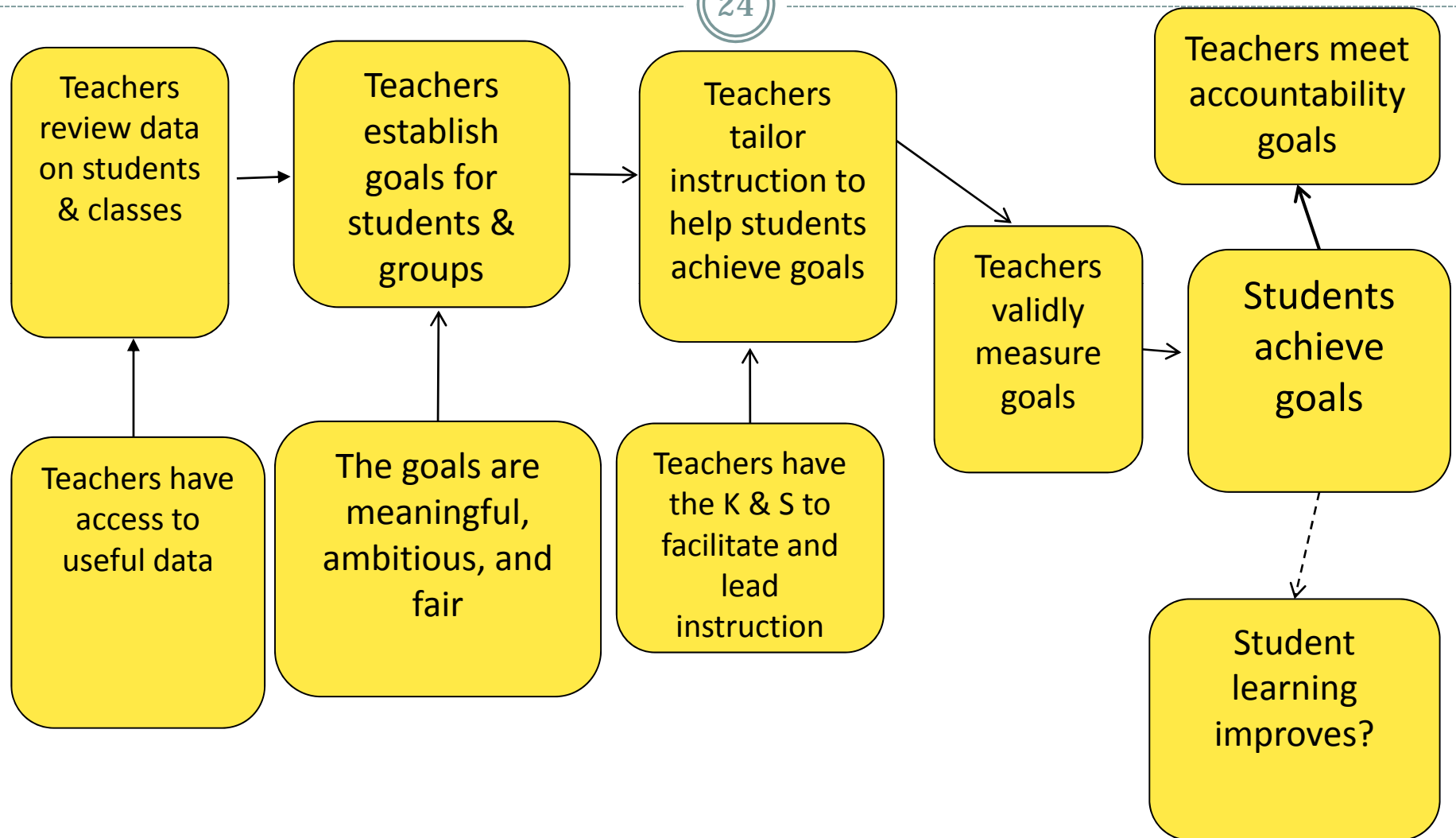
23

- An alternative approach is the “goal setting” approach used in several places (e.g., Denver Public Schools) and advocated by organizations such as the New Teachers Project.
- The basics....
  - Teachers, along with principals (or others), establish goals for each student or the class as a whole
  - A measurement or evaluation process is established to determine the extent to which teachers meet their goals
  - Teachers are then judged according to this measurement/evaluation process



# Theory of Action for Goal Setting Approach

24





# Claims, challenges, opportunities

25

- **Claim**: *Teachers have the knowledge, skills, and attitudes (& ethics) to set meaningful, ambitious, and fair goals for individual students*
- **Challenge**: Who will guide, monitor, and/or evaluate the quality of these goals?
  - This adds an extra (or at least different) significant validation requirement beyond what we saw with the previous approach.
- **Opportunity**: Teaching quality would like improve if teachers were supported in improving the way they used data to establish goals for individual students.



# More claims, challenges, opportunities

26

- Claim: *Teachers have the knowledge and skills to tailor learning opportunities for individual students*
- Challenge: Will there be a temptation to limit the range/variability of the goals to maximize efficiency?
- Opportunity: If teachers were really expected to focus on the needs of individual students, learning opportunities could very well improve. Would using group instead of individual goals limit this opportunity?



# And more claims, challenges and opportunities

27

- Claim: *Teachers and/or others have measurement or evaluation procedures sufficient for judging whether students have reached the intended goals*
- Challenge 1: Are classroom assessment tools capable of validly measuring ambitious goals?
- Challenge 2: If external assessments are used, would that lead to narrow goals to match the more limited tools (tail wagging the dog)?
- Opportunity: Could this be a lever for improving the quality of classroom assessment and evaluation tools and processes?



# The Theory of Action

28

- A helpful tool that quickly reveals where differences in the use context might cause a conflict when trying to implement a challenging policy
- Can help explore some alternative approaches for satisfying the policy goal of using local assessment information to contribute to the evaluation of educators involved with typically non-tested grades and subjects.
- More complete theories of action should be developed to help evaluate the logic and coherence of policies before moving to the implementation phase



# Loosen the “noose”

29

- All of these examples tried to work within the existing policy/legal requirements
- Using a theory of action or related approach may help reveal irreconcilable issues that can only be addressed by changing the use requirements (i.e., policy)
- The next steps would involve offering a more coherent use case as a foundation for a revised policy
  - But that’s the next paper....

