



Session 3 (Part 1)

Consequential Uses of Assessment: A Friendly Debate

Reidy Interactive Learning Series (RILS) Conference
Portsmouth, NH, September 26-27, 2024
AC Marriott Hotel

Access RILS Resources at:

<https://sites.google.com/nciea.org/rils2024>



The screenshot shows the homepage for the Reidy Interactive Learning Series (RILS) 2024. The header includes the RILS 2024 logo and navigation links for Home, About, Agenda, Location, and Presenters. The main content area features the title "The Reidy Interactive Learning Series (RILS)" and the subtitle "Consequential Uses of Assessment: Taking Stock and Looking Ahead". The event dates and location are listed as "September 26-27, 2024 Portsmouth, New Hampshire". Below this, there are four image-based links: "About RILS" (showing the Center for Assessment building), "Agenda and Resources" (showing hands writing on a notepad), "Presenters" (showing a group of people in a meeting), and "Location" (showing a building at night).



Overview (Part 1)

Proposition 1



Proposition 2



- Opening statement in favor of proposition (5 min)
- Opening statement opposed to the proposition (5 min)
- Moderator questions to team in favor of the proposition (7 min)
- Moderator questions to team opposed to the proposition (7 min)
- Closing statement from team opposed to the proposition (3 min)
- Closing statement from team in favor of the proposition (3 min)

Debaters



Scott Marion



E. Caroline Wylie

Team 1



Carla Evans



Brian Gong

Team 2

Debate

Proposition 1

Policies that require the use of large-scale assessments for consequential decisions for **students** (e.g., diploma eligibility, promotion/ retention) do more harm than good.

Arguing in Favor

Carla Evans & Brian Gong

*Policies that require the use of large-scale assessments for consequential decisions for students **do more harm than good.***

Why?

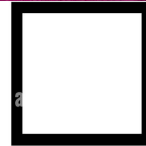
Large-scale policies in education, health care, law, etc., are blunt instruments.

The policies are going to be wrong for some individuals some of the time.

Two types of error:

1. “False alarm” (False Positive)
2. “Missed detection” (False Negative)

Three Example Policies



pass



fail

If Test Sent a False Alarm



Missed Detection



RILS High School

This certifies that

Scott Marion

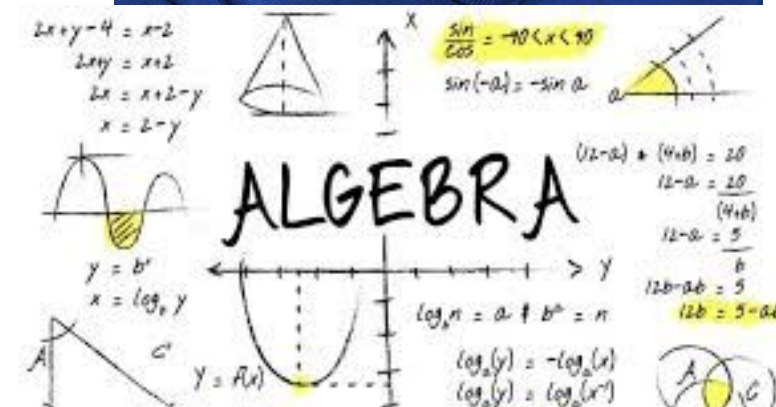
Has satisfactorily completed the necessary requirements of study as prescribed by school administrators and is thereby presented with this

HIGH SCHOOL DIPLOMA

And is entitled to all the rights and privileges pertaining thereto.

Pass

Fail



*In each of these circumstances, **is it better to lean on the side of a false alarm** and not allow Scott, Caroline, or Chris these opportunities because of a single test score or **is it better to err on the side of a missed detection** and ‘doing no harm’?*

*For these reasons, policies that require the singular use of large-scale assessments for **consequential decisions for individual students**—and ignore error, other data, and the full range of consequences— **do more harm than good.***

Arguing Against

Scott Marion & Caroline Wylie

Summary

- Students benefit from having skin in the game.
- Students MUST be supported by opportunity to learn and with interventions if they don't pass the test.
 1. Large-scale assessments help establish common expectations for students.
 1. Consequences help focus students' and teachers' attention on these expectations.
 2. Large-scale tests help us target the interventions to the students most in need.

Proposition 2

Policies that require the use of large-scale assessments for consequential decisions for **schools** (e.g., test results as prominent components of school accountability designations) do more harm than good.

Arguing in Favor

Scott Marion & Caroline Wylie

Summary

- Monitoring school quality is a responsible action to maintain public trust in the educational system.
- However, using test scores as the primary driver in this process is **irresponsible** and **misleading**, with limited evidence that it has actually made a difference for student learning.
- We have three concerns about how accountability systems are structured:
 1. The reliance on test scores may result in undue focus on the tested subjects.
 2. Relying on test scores and other outcome measures ignores critical input factors that often differ across schools even within the same district.
 3. A focus on testing to drive school accountability ignores the shared responsibility of other components of the system such as the responsibility of school districts to differentially support their schools.

Arguing Against

Carla Evans & Brian Gong

Summary: Con

- Policies that require the use of large-scale assessments for consequential decisions for **schools** (e.g., test results as prominent components of school accountability designations) **do not** do more harm than good.

Accountability, assessments are necessary

- Accountability for individuals and educational organizations is a wide-spread, accepted **principle**, but often admittedly, not as effective as wished for (there is a balanced between internal and external accountability/responsibility, across multiple actors in educational system)
- Assessing and assessments are **necessary** for feedback and evaluation, essential components of accountability and improvement
- **Large-scale assessments** are necessary when assessing is done at a scale where individual assessments are impractical and/or where standardization in instrumentation, administration, and scoring is highly valued

Assessment & accountability quality depend on purpose, design, implementation

- We're not arguing that every assessment is valid or every accountability system is helpful. A Center main contribution is helping states, districts, and others design and implement better policies, programs, and procedures, and problem-solve when there are areas to improve.
 - Ask us about accountability systems that can comply with federal requirements and also reflect other values (e.g., localization of indicators)!

Specific considerations 1 & 2

- Federal (and most states') theory of action: identify lowest performing schools in core content areas to target assistance
 - Schools **unlikely to improve** without outside intervention and support
 - Identification systems built to give states **legally defensible** means to declare districts/schools qualify for state intervention (e.g., comparable)
- Do we have the wrong indicators to identify very low performing schools in core content areas?
 - Title 1 (economic disadvantaged) focus on ELA/reading and math, not total school quality; hence not primarily to distinguish good from better schools
 - In many states, bottom 5% of schools would change marginally by reweighting indicators, including Growth

Specific consideration 3

- Claim: Identifying schools hasn't helped them appreciably get better
 - Several states have successful track records in helping schools identified for assistance improve performance enough to exit
- To get better at learning/teaching:



- I submit that ~~identification~~ of schools that need help has been largely ok (i.e., assessment and accountability systems have done their job). However, **Diagnosis** of root causes may be weak. **Prescription** of what should be done is often weak, especially for systemic barriers. And **Implementation** is rarely thorough. This is true at the state, district, school, and classroom levels too often.
 - E.g., Who has successfully implemented programs to help struggling readers in Grade 4 catch up? Who has implemented programs to retain high quality teachers in neediest schools?
 - Poor instructional policies and practices take root when people don't know how and/or don't have the resources to be successful using good instructional practices (e.g., narrowing the curriculum; cheating)—not because there is an accountability system. Very few people cheat on state accountability tests.

Break



Session 3 (Part 1)

Consequential Uses of Assessment: A Friendly Debate

Reidy Interactive Learning Series (RILS) Conference
Portsmouth, NH, September 26-27, 2024
AC Marriott Hotel