

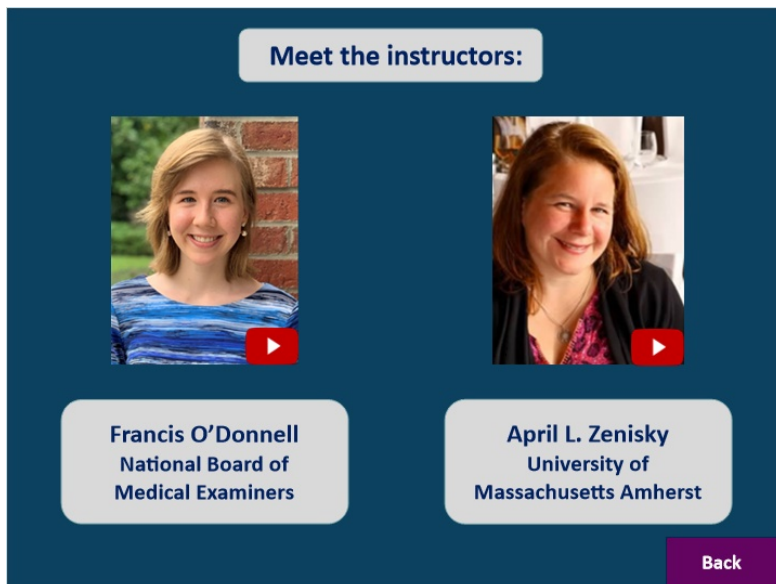
DM21 SLIDES (Results Reporting, Version 1.0)

1. Module Overview

1.1 Module Cover (START)




1.2 Instructors




1.3 Designers

Meet the designers:




Xi Lu
Florida State
University



André A. Rupp
Mindful
Measurement

Back

1.4 Welcome



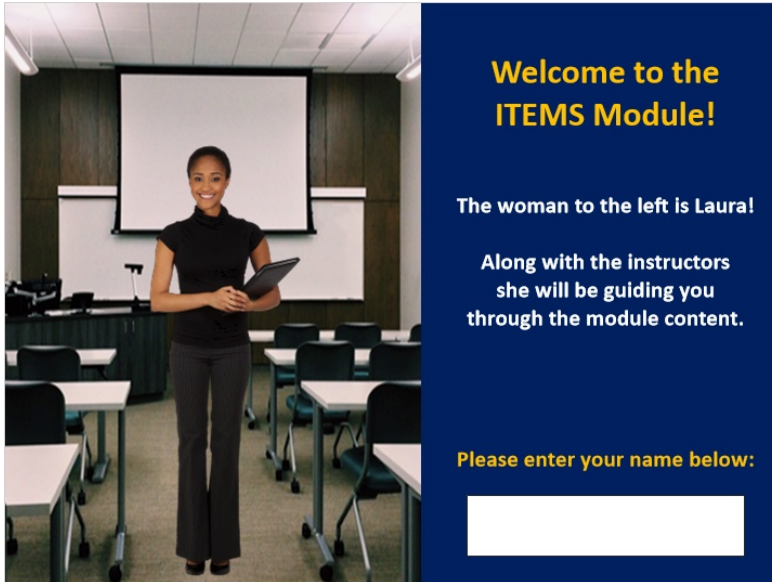
**Welcome to the
ITEMS Module!**

The woman to the left is Laura!

Along with the instructors
she will be guiding you
through the module content.

Please enter your name below:

Untitled Layer 1 (Slide Layer)



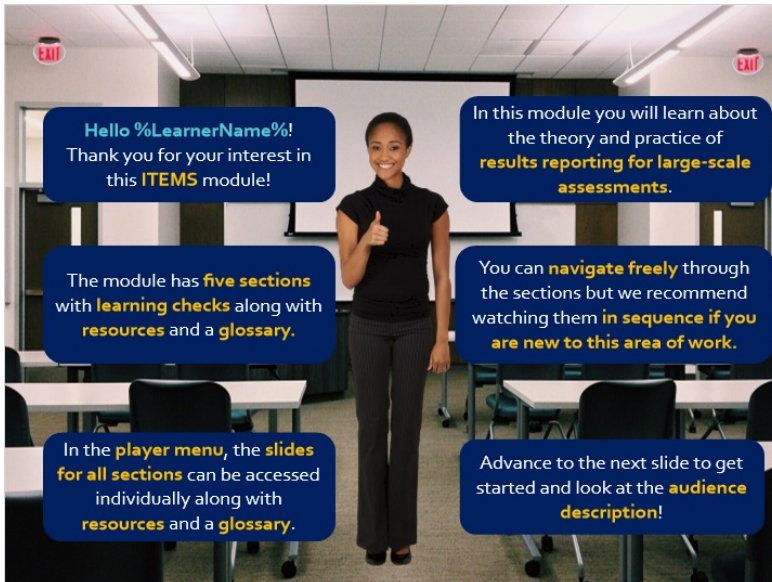
**Welcome to the
ITEMS Module!**

The woman to the left is Laura!

Along with the instructors
she will be guiding you
through the module content.

Please enter your name below:

1.5 Overview



Hello %LearnerName%!
Thank you for your interest in
this **ITEMS** module!

The module has **five sections**
with **learning checks** along with
resources and a **glossary**.

In the **player menu**, the **slides**
for all sections can be accessed
individually along with
resources and a **glossary**.

In this module you will learn about
the theory and practice of
**results reporting for large-scale
assessments**.

You can **navigate freely** through
the sections but we recommend
watching them **in sequence if you
are new to this area of work**.

Advance to the next slide to get
started and look at the **audience
description!**

1.6 Target Audience

Target Audience

Anyone who would like a gentle but methodologically sound introduction to this topic:

- graduate students and faculty in Master's, Ph.D., or certificate programs
- psychometricians and other measurement professionals
- data scientists / analysts
- research assistants or research scientists
- technical project directors
- assessment developers



However, we hope that you find the information in this module useful no matter what your official title or role in an organization is!

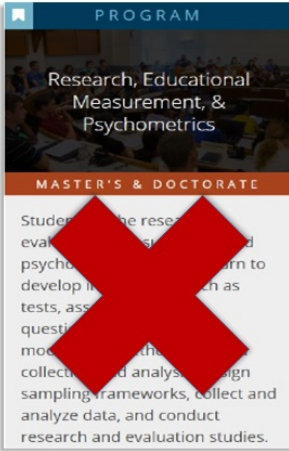

1.7 Expectations (I)



Let's discuss expectations....

1.8 Expectations (II)

ITEMS Modules in Context



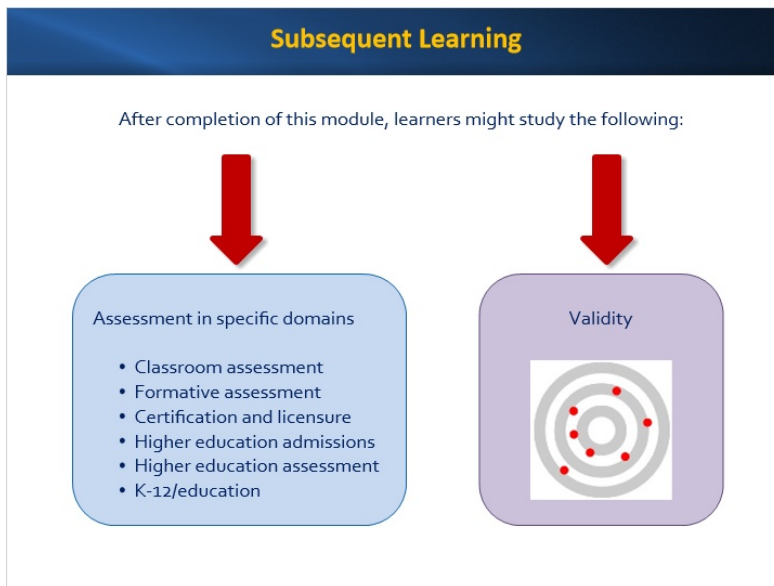
1.9 Learning Objectives

Learning Objectives

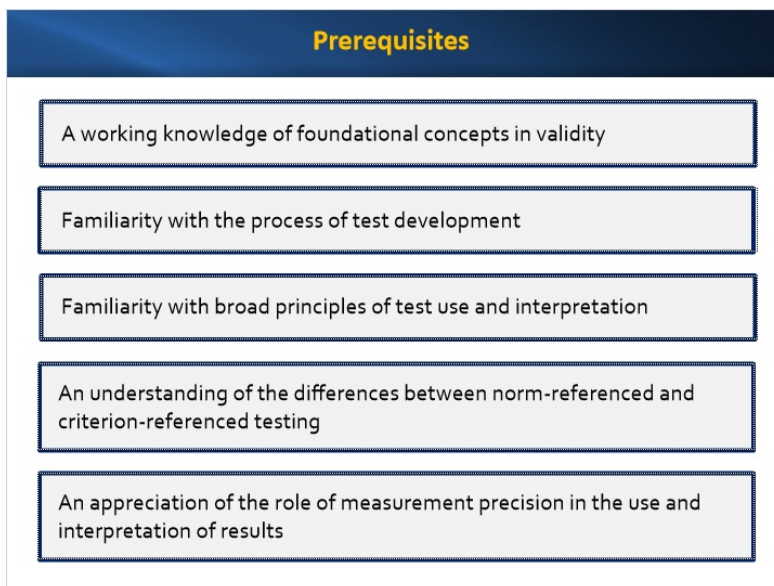


1. Articulate how results reporting connects to validity arguments
2. Understand the broad principles and phases of report development
3. Think critically about the effectiveness of different kinds and presentations of information relative to report users' needs
4. Articulate the core ideas of three approaches for conceptualizing research on results reporting
5. Apply a working knowledge of empirical findings on report elements, layout, and appearance from the literature
6. Develop a working understanding of how to apply an established model for the development, evaluation, and maintenance of results reports

1.10 Subsequent Learning



1.11 Prerequisites




1.12 Module Citation & Resource



Module Citation & Resource

Module Citation

O'Donnell, F., & Zenisky, A. L. (2020). Score reporting for large-scale assessments (Digital ITEMS Module 21). *Educational Measurement: Issues and Practice*, 39(4), 138-139.

Resource

1.13 Main Menu

  **Main Menu**

01

Introduction and Overview
[20 Minutes]

02

The Who, What, and How of Results Reporting
[20 Minutes]

03

Report Development – Modeling the Process
[15 Minutes]

04

Best Practices and Guidelines for Reporting Results
[20 Minutes]

05

Conceptualizing Research on Results Reporting
[20 Minutes]

2. Section 1: Introduction & Overview

2.1 Cover: Section 1



2.2 Objectives: Section 1

Learning Objectives




I. Define major ways of describing test performance

III. Discuss how user needs and wants influence report development


II. Explain recent shifts in report development and use

IV. Describe the concept of reports as data stories

2.3 Topic Selection



Introduction and Overview



Shifts in Report Design and Use



Reports Deconstructed



Reports as Data Stories

Click on each button to learn more



End Section

2.4 Bookmark: A Seismic Shift





Shifts in Report Design and Use

2.5 A Seismic Shift I



A Seismic Shift




- Reporting is about **communication of performance**
- Performance on any test can be **represented in multiple ways**

2.6 More about Results Reporting





More about Results Reporting

- Most “score reports” produced in large-scale testing today provide **data about performance** that includes scores
- They also use other strategies to **represent and characterize** examinee knowledge and skills
- Many of these other results are **derivations of scores**




2.7 Ways of Describing Test Performance



Ways of Describing Test Performance

- ✓ Total scores
- ✓ Performance / achievement levels
- ✓ Percentile ranks
- ✓ Subdomain scores
- ✓ Subdomain performance levels
- ✓ Item-level performance
- ✓ Growth scores
- ✓ Other



Some things are scores,
but not everything!

2.8 Score reporting



Score Reporting

'Score reporting' today is about more than a score.
Focus increasingly is on reports that **tell a more complete data-driven story**
about how examinees did on the assessment or learning activity

-  1. Typically start with overall performance
-  2. Then move to a smaller grain size
-  3. Represent performance in multiple ways
-  4. Then use the presented results to identify potential actions

2.9 A Seismic Shift I



A Seismic Shift



Score reporting is an important **starting place**.



Reports are **data stories**
as any given report serves a specific purpose


- different contexts for testing
- different audiences
- different uses of tests
- communicate certain, specific information
- achieve a particular aim for a particular user group

2.10 Topic Selection




Methods to Unfold Data Stories

Method 1




Standards

Method 2




Audiences

Method 3



Report

Click on each button to learn more.



2.11 Standards Overview



Standards 6. 10 Review

When test score information is released, those responsible for testing programs should provide interpretations appropriate to the audience. The interpretations should describe in simple language **what the test covers**, **what scores represent**, **the precision/reliability of the scores**, and **how scores are intended to be used**.

(Standard 6.10; AERA, APA, & NCME, 2014)



2.12 Promoting Meaning





Promoting Meaning

- Audience matters
- Define the domain
- Context, context, context
- Results beyond numbers
(There's a range!)
- It's all about intent
(Validity!)



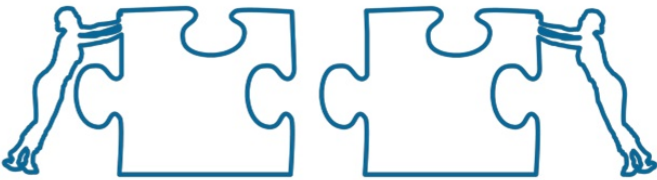
2.13 Context for Meaning





Context


- Enhancing context helps to clarify meaning
- Better understanding of meaning improves relevance

Greater relevance = greater likelihood of action, use, and/or value



2.14 Bookend: Standard








This is the end of this part.

Method Selection

2.15 Report Overview




What's a Report?



Document generated by agency
charged with reporting and
disseminated to users intact

"Here's where to look, what you'll
see, and prepped answers to
questions you'll probably ask"

Agency-generated
Explanatory and Instructive




'Document' generated by user
from agency database to be explored
as interest guides them

"Go forth and customize which results
will display based on your interests
and needs"

User-generated
Explanatory and Constructive

2.16 Reflection




Reflection

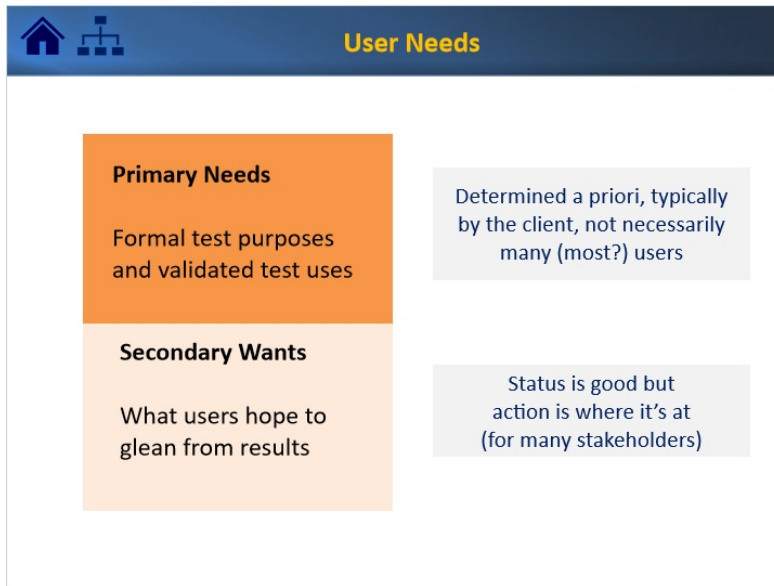
- What make a report "effective" for different users, today?
- What makes a report "not effective", today?
- What does "effective" even mean?

Girl?

Fish in the ocean?



2.17 User needs



2.18 Personal Reflection #1



The slide is titled "Learning Check: Personal Reflection #1" and features a header with a home icon and a tree icon. On the left, there is a photo of a woman in a black top and pants, standing with her hand on her chin in a thinking pose. To the right of the photo is a text box with instructions for a reflection exercise.

Consider a testing program that you work on or that you are familiar with.

For a **current report** that is provided to an **audience of your choosing** for the testing program you have in mind:

- **Define** whether the current report is **static or interactively generated** by the user.
- **Jot down a reason or two** why that report format is **appropriate for that audience** OR **describe in brief** how a shift to the other format might have **benefits for the audience's use or understanding** of the results.



2.19 Boundaries




Generalization

- **Primary focus of module is on large-scale assessment**
 - ✓ K-12
 - ✓ Certification and licensure
 - ✓ Higher education admissions
- **Ideas are not test-specific**
 - ✓ The choices made in developing reports impact what users see
 - ✓ And ultimately, what users take away from tests and test results
- **Some choices are better than others**

2.20 Bookend: Report

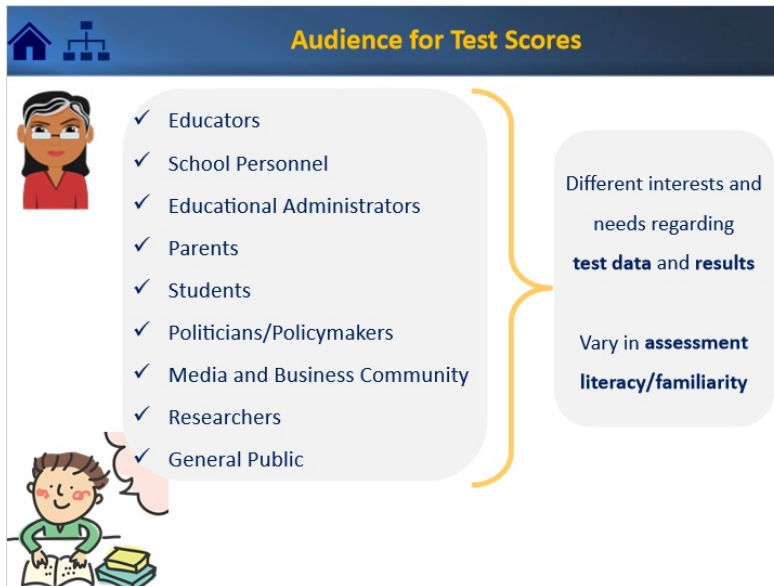




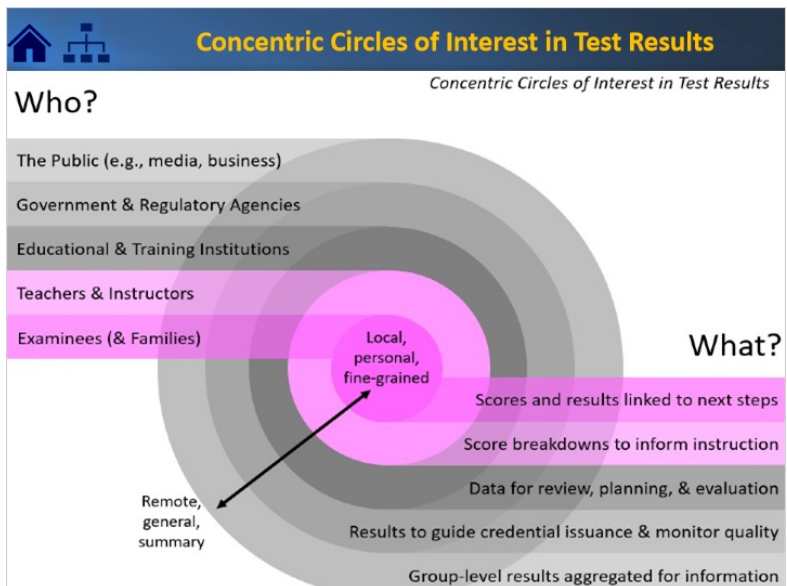
This is the end of this part.

Topic Selection



2.21 Audience Overview



2.22 Circles of Interest




2.23 Prioritize Intended Users






Prioritize Intended Users

- Identify **main use cases**
- Develop reports to **accomplish known tasks**
- Where feasible, develop systems to **aid users with task-specific data analysis needs**
- Provide supports to **address known issues** for **various user groups**



2.24 Bookend: Audience





This is the end of this part.

Method Selection

2.25 Bookmark: Deconstructing a Report



2.26 What is a Report?



What is a Report?



What's the image that comes to mind?

What would you expect to see on a report ...

- ...in the K-12 context...
- ...for a summative test...
- ...for an individual student?



2.27 Deconstructing a report II





Deconstructing a Report

Let us take a brief and scenic walk through those elements, with a fictional but illustrative report



2.28 Topic Selection




Topic Selection

Report as a "document" composed of **specific** informational elements, **purposefully** included and arranged in a **deliberate** sequence
(inspired by Hattie, 2009)

Introductory content

Finer-grained results




1 2 3 4



High-level results

Conclusions & next steps

Click on each numbered circle to learn more.




2.29 Introductory Content (I)



Introductory Content (I)

Name: Sofia Student School: Fictional Elementary
Grade: 5 District: Fictional District
Subject: English Language Arts (ELA)



SOFIA'S ENGLISH LANGUAGE ARTS RESULTS

Sofia took an assessment measuring her knowledge of Fictional State's English Language Arts learning standards. The standards reflect what knowledge and skills she will need to succeed in this subject next year.

Personalization/individualization – locate individual in context:
Age and physical space

Answer key questions:
What was assessed? To what end?

2.30 Introductory Content (II)

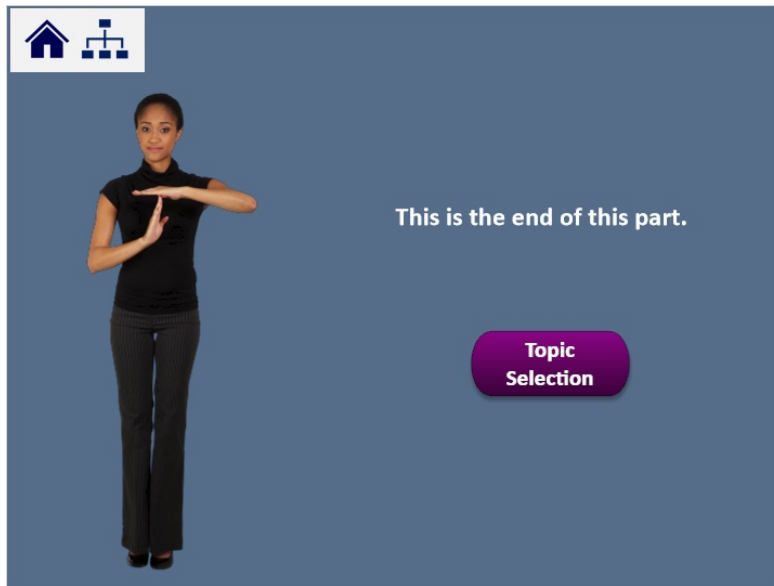


Introductory Content (II)

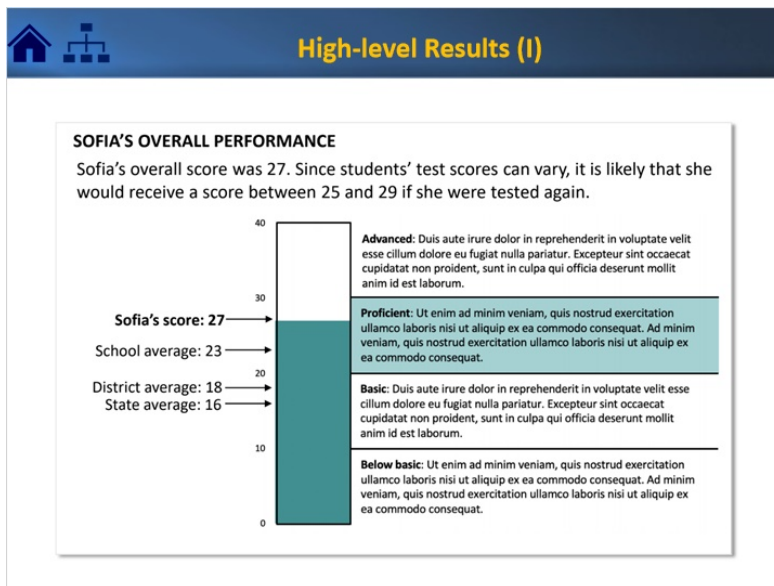
Sometimes agencies will include **specific introductory elements**.

- Official imagery (state seal or organization logo)
- Introductory letter from authority figure (commissioner/superintendent, executive officer)
- Text describing purpose of assessment
- Text describing purpose of report
- Information about content of assessment



2.31 Bookend: Introductory Content



2.32 High-level Results (I)



2.33 High-level Results (II)

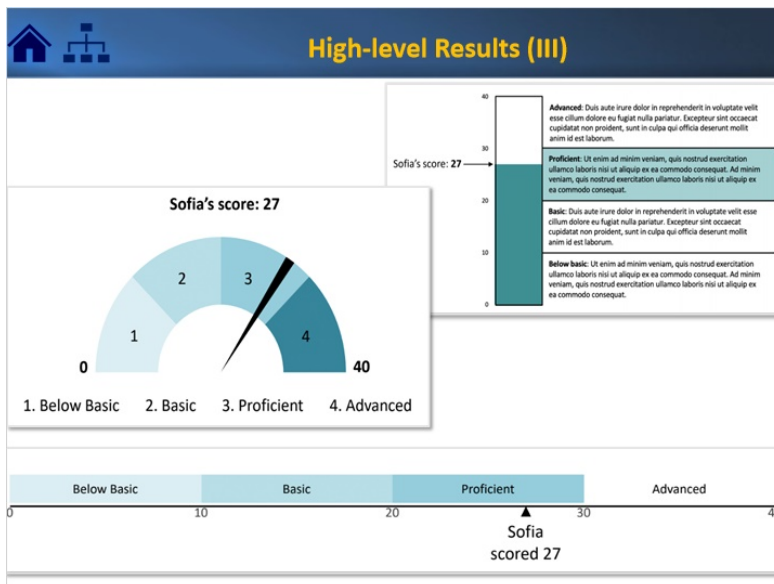


High-level Results (II)



Different visualizations and texts **highlight different aspects**.

- Provide overall results for multiple content areas in one image
- Other graphical strategies (bars, lines)
- Text and/or tables
- Manipulate font size, color

2.34 High-level Results (III)



2.35 High-level Results (IV)

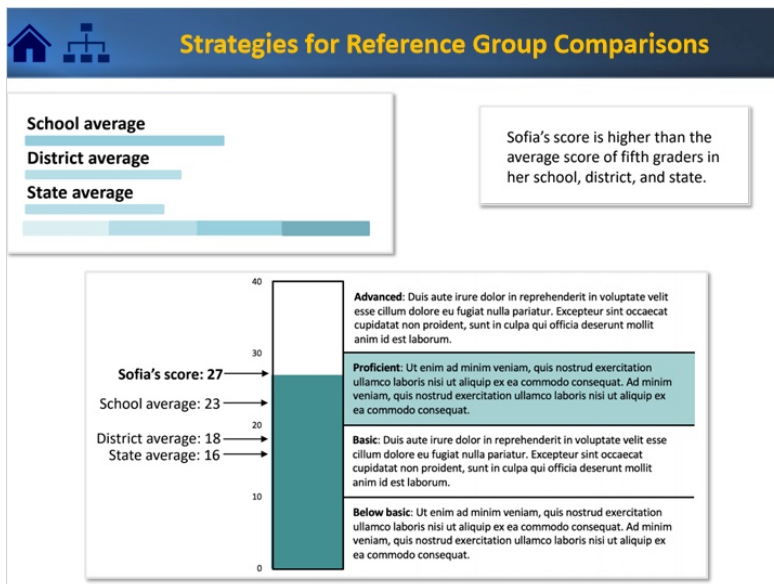


Reference Group Comparisons

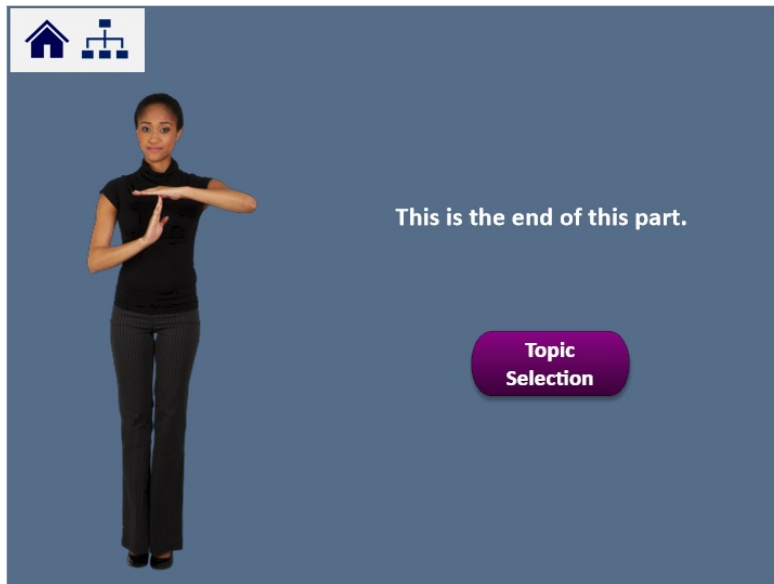
Provide **normative context** for the individual's performance.

- Using reference groups that may be salient to users
- School, district, state (others?)
- Not about Sofia's knowledge relative to content
- Positioning her performance relative to that of other 5th graders in some progression of groupings

2.36 High-level Results (V)



2.37 Bookend: Higher-level Results





2.38 Finer-grained Results (I)

Finer-grained Results (I)

In addition to overall scores, people often want to get as detailed as possible in reviewing results.




- Why? Why is that the grail of reporting?
- In a word: Improvement. And that's something that should be encouraged!
- Statistical elephant in the room – reliability of subscores
- Subscores are psychometrically challenging and this is an active area of psychometric research
- Agencies employ many strategies to provide subscore information to audiences

2.39 Finer-grained Results (II)

Finer-grained Results (II)

How did Sofia do on the different areas of the test?

| Area | Result | Description |
|--------------------|---|--|
| Comprehension |  | Sofia can effectively iaculis eu diam phasellus lorem. |
| Literary Text |  | Sofia may need support to eget vel risus praesent. |
| Informational Text |  | Sofia may be able to nec tincidunt praesent feugiat. |

Performance by Domain

Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in. Mattis vulputate enim nulla aliquet. Amet tellus cras enim.

| Content Domains | Sofia's Score | Score at Passing | Strength or Area of Need |
|--------------------|---------------|------------------|--------------------------|
| Comprehension | 9 | 8 | 1 |
| Literary Text | 6 | 7 | -1 |
| Informational Text | 7 | 7 | 0 |



Subscores

Sofia's test performance

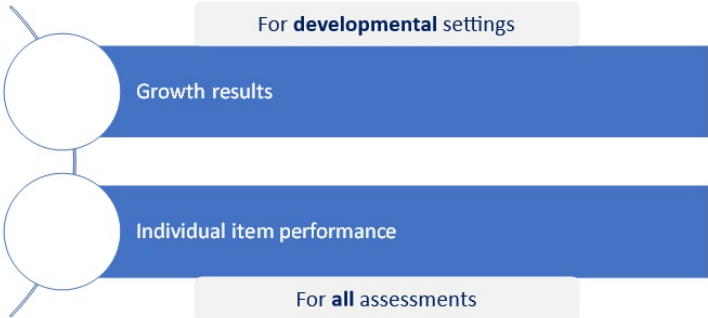
Explanation of how to use this information. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in.

| | Percent of points earned | Proficiency |
|--|--------------------------|-------------|
| Comprehension Content domain description. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris. | 87% | 61-82% |
| Literary Text Content domain description. Feugiat vivamus at augue eget arcu dictum varius duis at. Amet justo donec enim diam vulputate ut pharetra sit amet. | 61% | 65-79% |
| Informational Text Content domain description. Mattis vulputate enim nulla aliquet. Amet tellus cras adipiscing enim. | 70% | 65-79% |

2.40 Finer-grained Results (III)

Finer-grained Results (III)



For **developmental** settings

Growth results

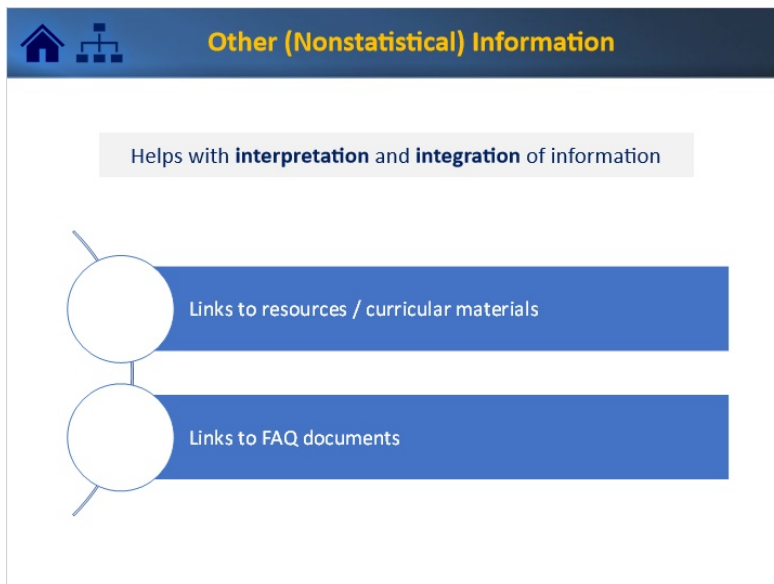
Individual item performance

For **all** assessments

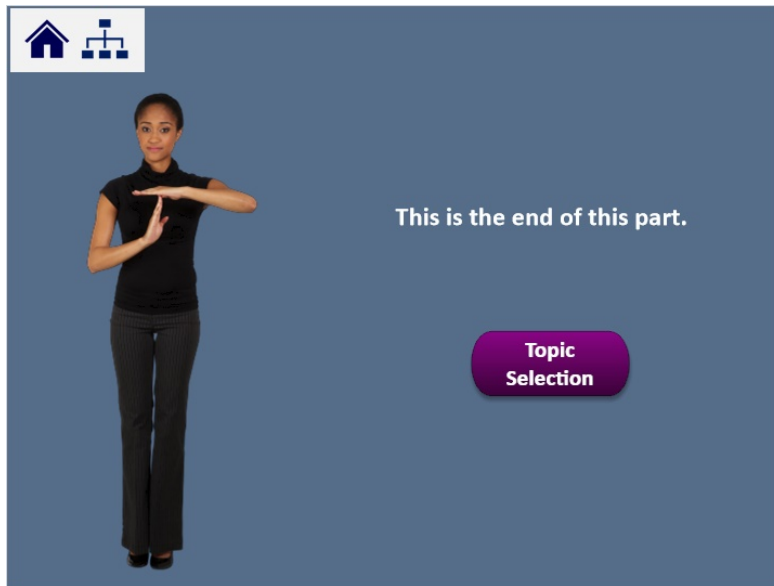
2.41 Bookend: Finer-grained Results



2.42 Other Information





2.43 Bookmark: Other Information



2.44 Bookmark: Reports as Data Stories




2.45 Results Report as Data Story (I)



Results Report as Data Story (I)



A results report **tells a data story** for a particular audience, and **contains specific information**, and **is designed and formatted and delivered** to help accomplish something.

Data **Analysis** **Story**



| x | y |
|----|---|
| 3 | 4 |
| 7 | 5 |
| 5 | 6 |
| 10 | 7 |

2.46 Results Report as Data Story (II)



Results Report as Data Story (II)

Start with personalization and individualization of report

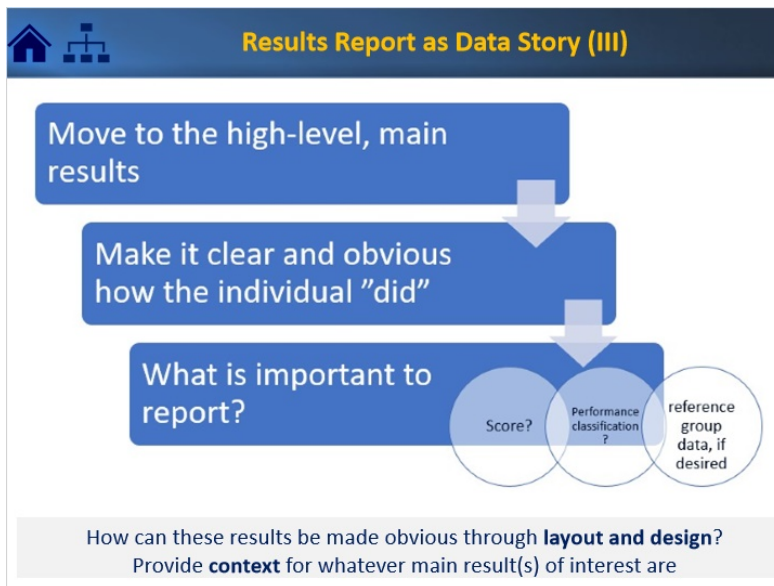
↓

Provide info about the report document and the assessment results being reported

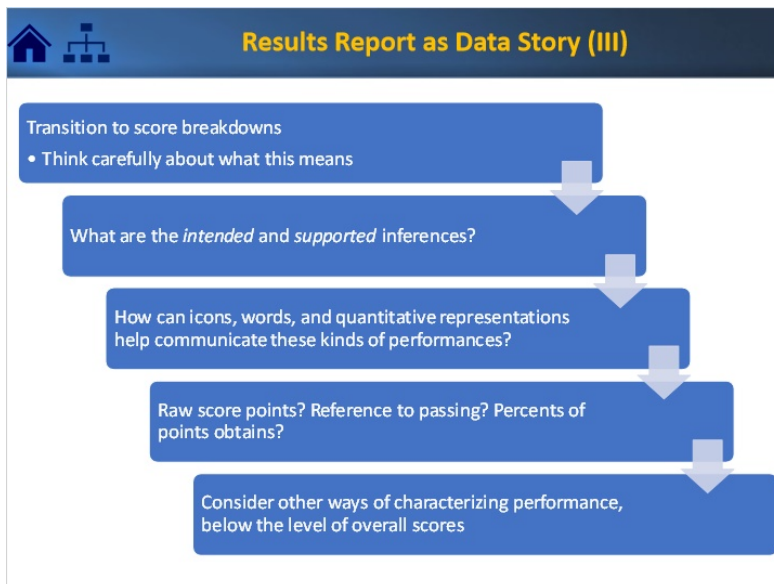
Draw in the intended user:

- ✓ Who is this document about?
- ✓ Why should I attend to it?
- ✓ What does it tell me?

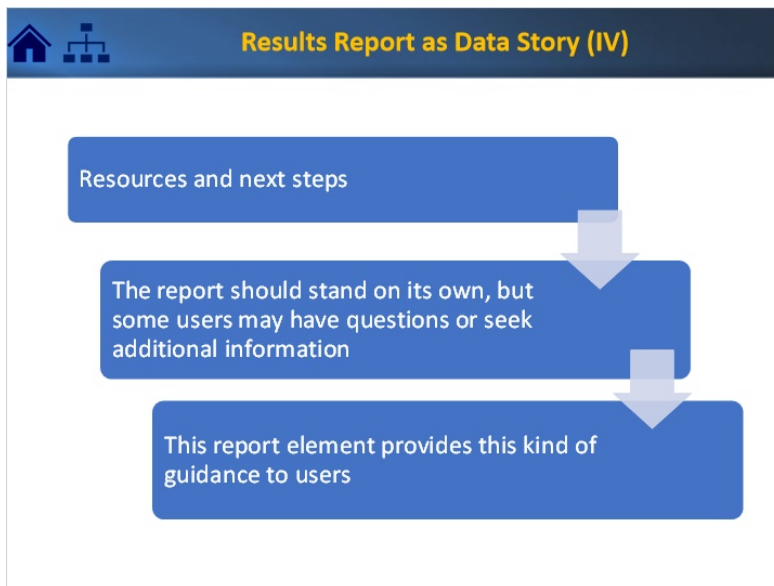
2.47 Results Report as Data Story (III)



2.48 Results Report as Data Story (III)



2.49 Results Report as Data Story (IV)





2.50 Results as Data Story (V)


The icon is an orange rounded square containing a white graphic of a right-pointing arrow with a curved tail that loops back to the start of the arrow.

- Report should **guarantee safe passage** (attend to important results then flow to details)
- The story of the report should **have a clear theme**
- A report should be designed to **answer specific questions**
- Reports should be conceived of as **actions** not a screen to print

2.51 Personal Reflection #2





Learning Check: Personal Reflection #2




Consider again a results report for a testing program you work on or are familiar with.

- **Jot down a quick list** of the **main information** conveyed in **each identifiable section** of the report.
- **As you look at your list** of info in each identifiable section of the report, what **reporting aim** does each section of the report accomplish?
- **How successful**, in your opinion, is each section in **accomplishing that aim**?
- **How might you decide** if each section accomplishes that reporting aim?

2.52 End of Section 1



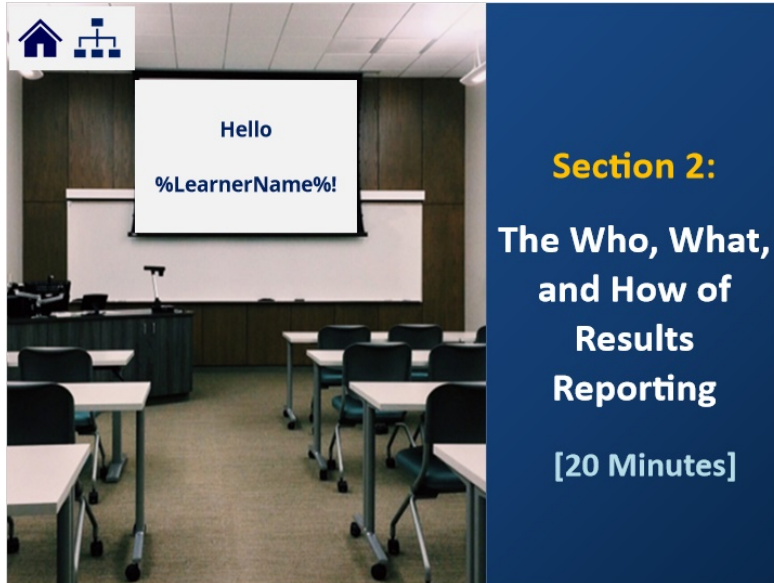


This is the end of this part.

Topic Selection

3. Section 2: The Who, What, and How of Results Reporting

3.1 Cover: Section 2




3.2 Objectives: Section 2

The image shows a slide titled "Learning Objectives" with a blue header. Below the header is a target icon with a red bullseye and a black arrow hitting the center. Below the target are three numbered objectives in separate boxes:

- I. Understand what user characteristics should be considered in report development
- II. Think critically about the effectiveness of different kinds and presentations of information
- III. Describe major design considerations affected by report delivery mode

3.3 Topic Selection



Who: Report users

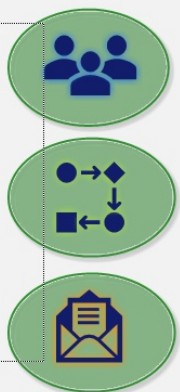
- audience considerations

What: Report components

- information design

How: Report delivery

- platform considerations



Learning Check

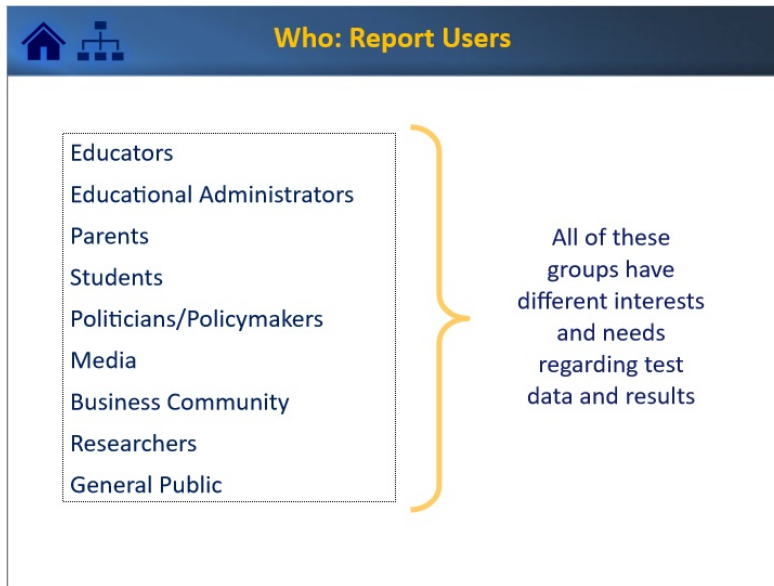
Section End

3.4 Bookmark:Who

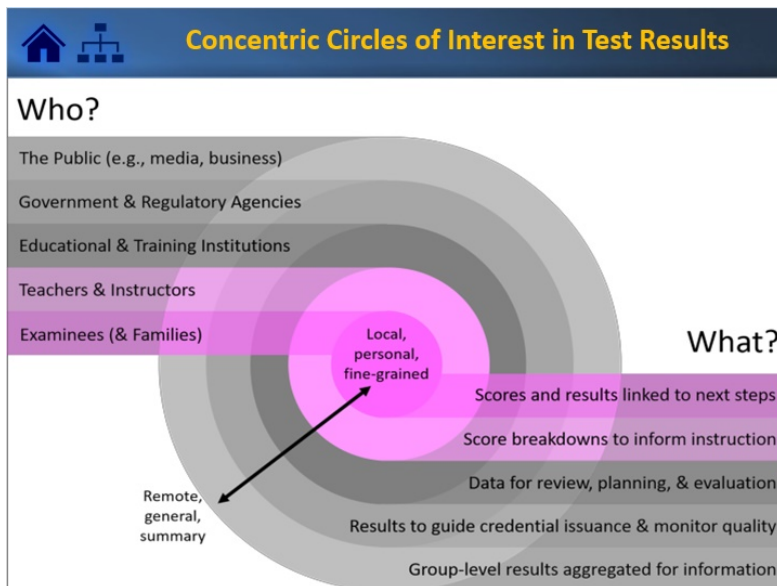


Who:
Audience
Considerations


3.5 Who I



3.6 Who II




3.7 Who III




Considering User Characteristics

Zapata-Rivera and Katz (2014) recommend conducting an **audience analysis** to connect the score report design process and report specifics to audience characteristics, noting:

“Although most design guidelines focus on making score reports understandable to people who are not testing professionals, audiences should be defined by more than just their lack of statistical knowledge.



3.8 Who IV



Audience Analysis (Zapata-Rivera & Katz, 2014)

Needs

- What are the users’ goals in using the report?
- What do they want to accomplish?

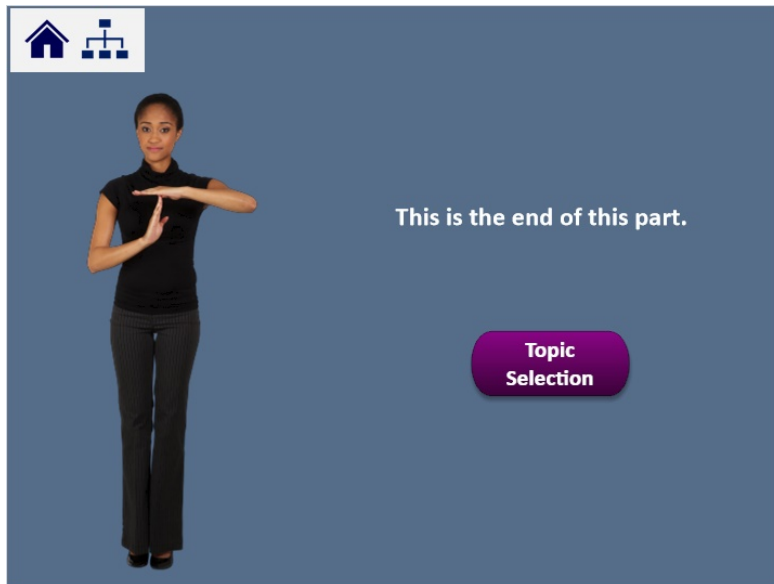
Knowledge

- What do users already know about the assessment and its context?
- What knowledge gaps might impact their understanding?

Attitudes

- What are users’ perceptions of the assessment?
- How might the design of the report contribute to a productive attitude about the results?

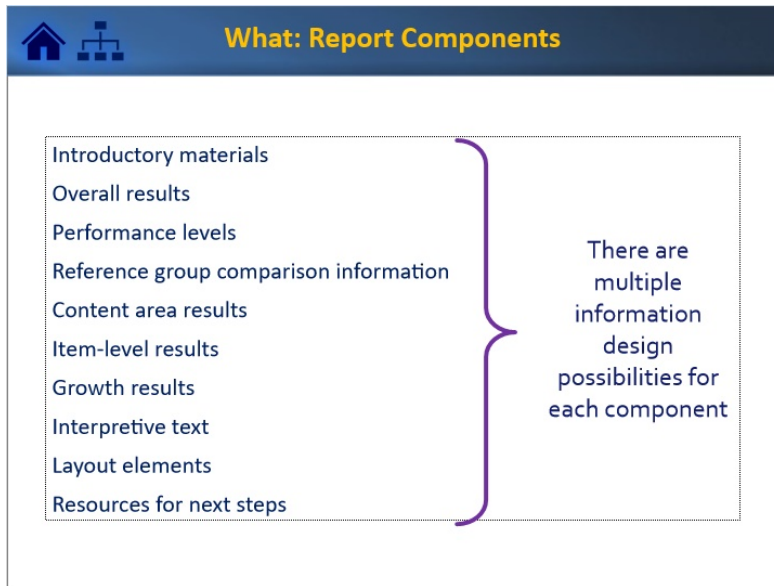
3.9 Bookend: Who



3.10 Bookmark: What



3.11 What I



3.12 What II

Report Component Examples

The following examples are from state-level departments of education and testing organizations that post score report samples online.

The samples are beneficial to the public and our community of practice.

We selected excerpts that illustrate a variety of approaches and removed identifying details from some.

Many thanks to all states and organizations that provide report samples, especially:

Arizona, California, Connecticut, Delaware, Georgia, Hawaii, Illinois, Indiana, Kansas, Massachusetts, Nevada, New Hampshire, New York, Rhode Island, South Carolina, Utah, West Virginia, ACT, and SAT

3.13 What III

Introductory Materials

Dear Parent/Guardian:

This report shows how Bella scored on the tests for English language arts/literacy and mathematics.

These online assessments were new this year for Bella as an eleventh-grade student. These tests are based on California's rigorous academic standards, which are designed to help every student graduate ready for college and a 21st-century career. Bella's scores provide an indication of her readiness for college-level work after graduation. Please see the Early Assessment Program portion of this report for information about steps Bella can take in twelfth grade to be ready for college.

While tests are just one way to measure Bella's progress, the results can help teachers and the school focus on areas in which students need more help. I encourage you to be involved in your child's learning and discuss these results with Bella's teacher(s).

Sincerely,

Sam Superintendent
State Superintendent of Public Instruction

Family Report

Jared A. Doe

Grade: 8
Date of Birth: 4/26/2003
Student ID: 123456789
School District: District (15)
School: Academy (12345)
Test Date:

What is in this report?

- Jared's ELA/Literacy and Mathematics scores
- Descriptions of Jared's performance on each claim/area of assessment
- FAQs and additional resources

About This Report

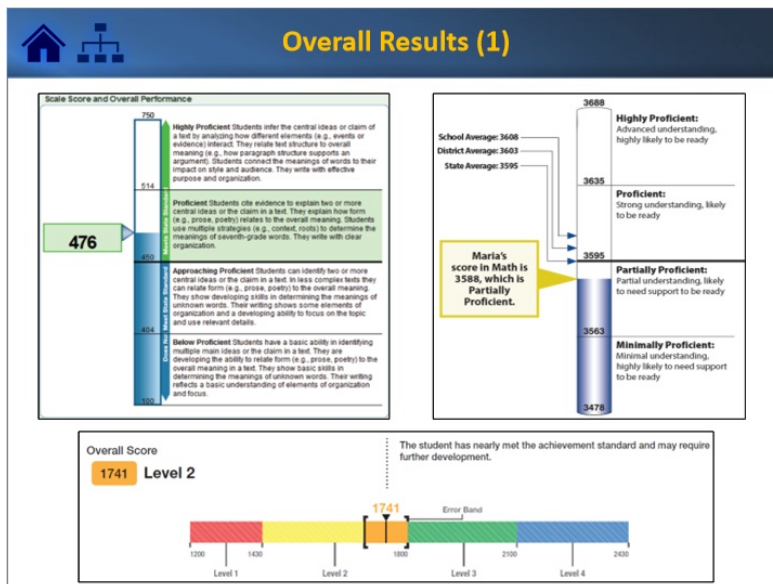
Front:

- Maria's overall score for this assessment includes a numeric score and a proficiency level.
- Her numeric score can be compared with the school, district, and state averages.
- The proficiency level indicates how well students understand current grade level material and how likely they are to be ready for the next grade.

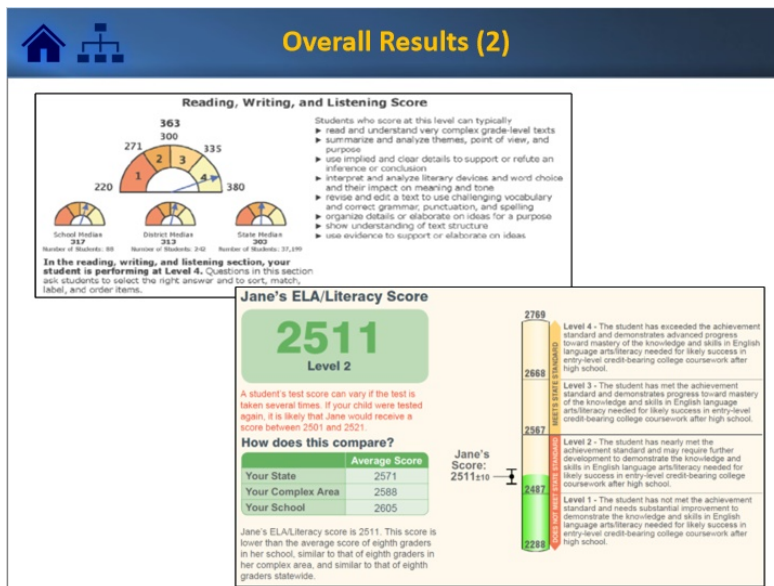
Back:

- Maria's level of mastery is shown for each scoring category.
- Scoring categories represent specific knowledge and skills included in this assessment.
- There is a detailed description of the mastery level for each scoring category.

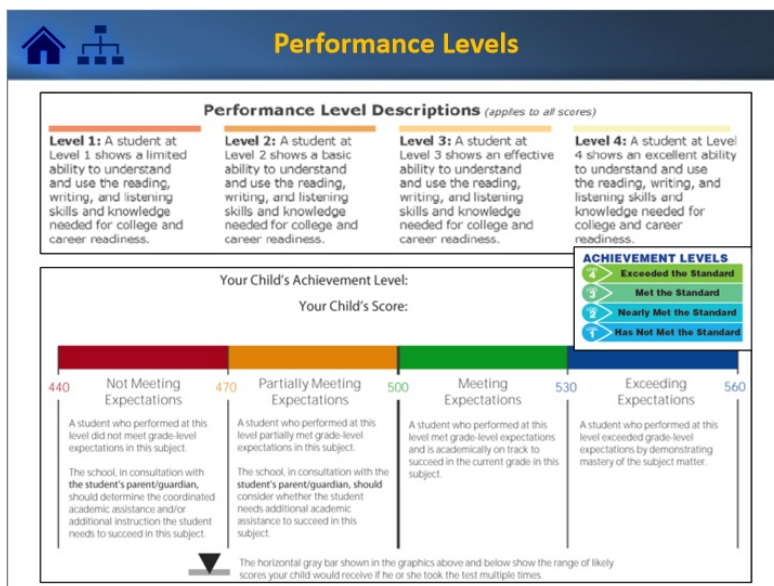
3.14 What IV



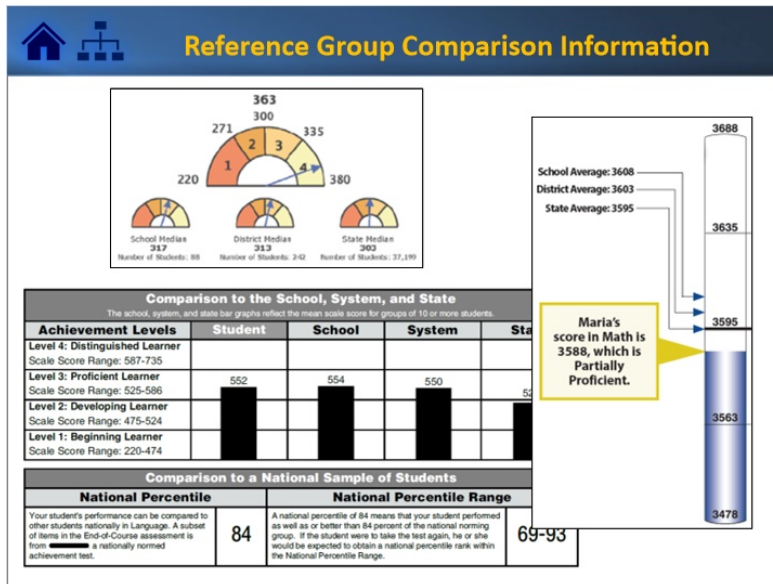
3.15 What V



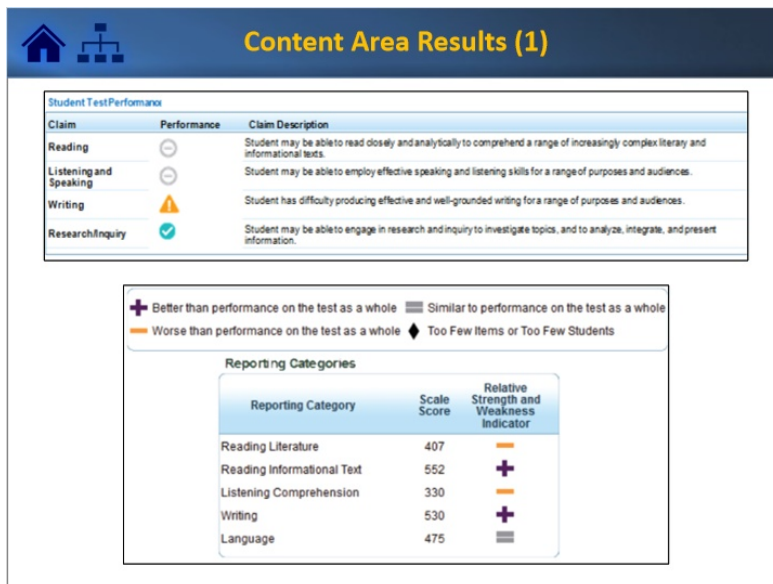
3.16 What VI



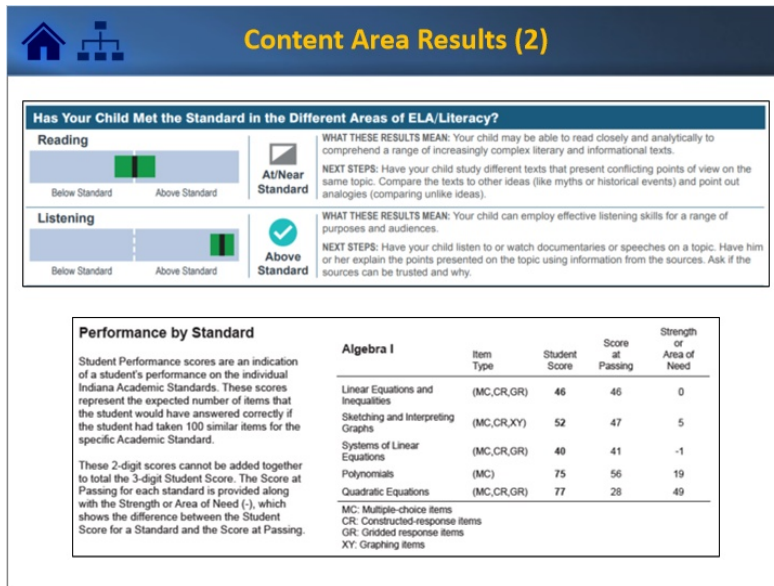
3.17 What VII



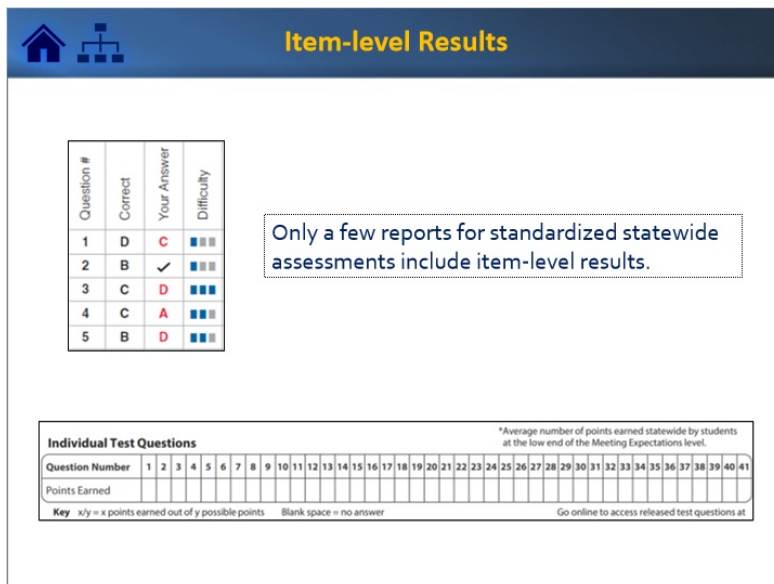
3.18 What VIII



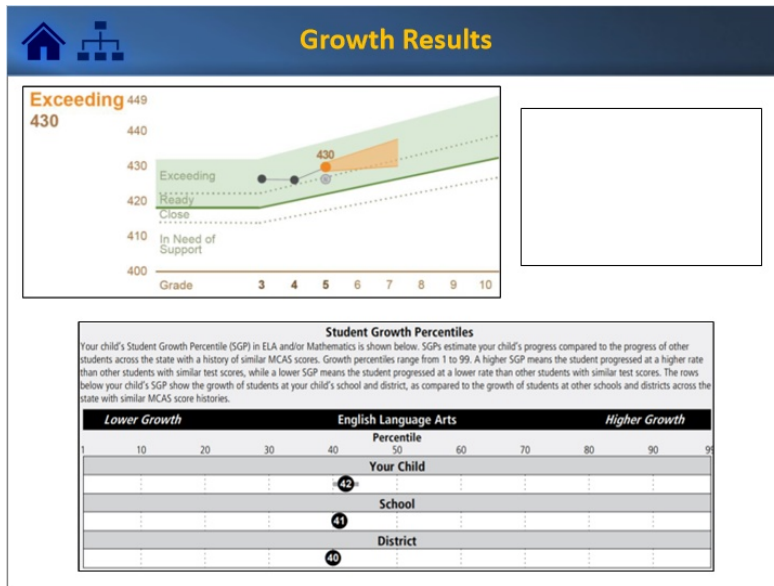
3.19 What IX



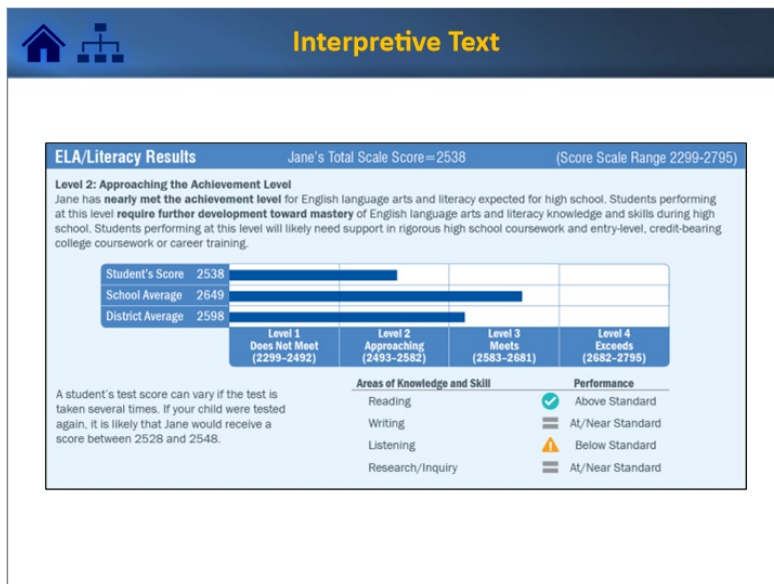
3.20 What X



3.21 What XI

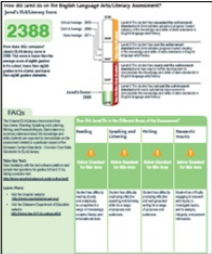
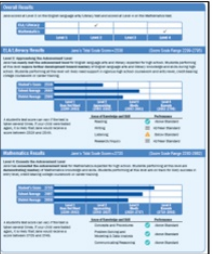



3.22 What XII



3.23 What XIII

Layout Elements

3.24 What XIV

Resources to Support Next Steps

SUPPORTING JANE'S SUCCESS

➤ Contact Jane's teacher and ask which skills are the most challenging for Jane.

➤ Ask your school how you will receive regular reports on Jane's progress.

➤ If appropriate, seek student support services such as counseling or school guidance to help Jane be most successful.

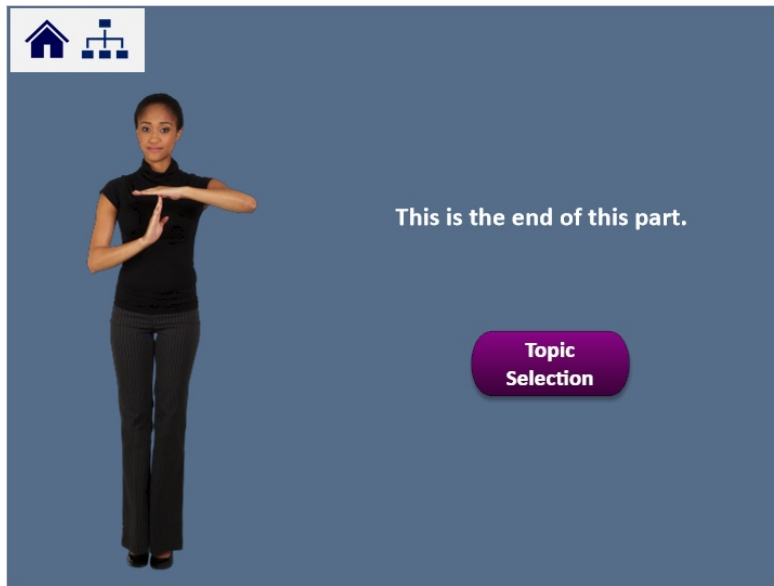
➤ Review classwork and homework with Jane to see how Jane's skills are progressing.

For more information about this test, the New York State standards, and how you can help Jane, go to:

| Your Student's Lexile Information | | |
|--|--------------------------------|---------------|
| Lexile Measure: 1315L Lexile Range: 1215L-1365L | | |
| <small>The Lexile Framework® for Reading matches a student's reading ability with the difficulty of text material. When selecting books, it is important to consider that Lexiles do not address age-appropriateness, student interest, or text quality. Suggested titles are not necessarily endorsed by the Department of Education. Books within the student's Lexile range can be found at the local library or by using the Find-a-Book database at www.lexile.com. For more information, visit www.gardens.org/lexile.aspx.</small> | | |
| Leisure Reading: 1215L-1315L | | |
| Suggested Titles | Author | Lexile |
| Hold Fast Your Dreams | Boyko, Carrie; Colen, Kimberly | 1220L |
| Don't Know Much about History | Davis, Kenneth C. | 1270L |
| The Quest for Mach One | Yeager, Chuck et. al. | 1310L |
| Motivating Challenge: 1315L-1365L | | |
| Suggested Titles | Author | Lexile |
| The House of the Seven Gables | Hawthorne, Nathaniel | 1320L |
| Walden | Thoreau, Henry David | 1340L |
| Robinson Crusoe | Defoe, Daniel | 1360L |

+ FAQ section, glossary, links to online documents



3.25 Bookend: What



3.26 Bookmark: How



3.27 How I

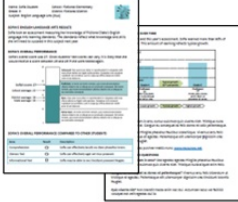


How: Report Delivery

For decades, the content of reports was bound by the margins of 8.5" by 11" sheets

- Usually 2-4 pages
- 1-2 double-sided sheets of paper



(Goodman & Hambleton, 2004; Rick & Park, 2017)



Advances in technology have made new options available:

- Static Online Reports
- Interactive Reports and Tools
- Informational Web Pages

3.28 How II





Static Reports

Static reports have been the focus of most results reporting studies and recommendations to date

All users in the same group see the same elements, with no interactivity

Whether static reports are delivered on paper or virtually, they may still be highly effective and innovative



3.29 How III

Interactive Reports and Tools

Interactive reports and tools are not yet the norm, but have been successfully implemented by a few organizations

Example: NAEP Data Explorer

11

From "Here's the Story" to "You're in Charge"

Developing and Maintaining Large-Scale Online Test- and Score-Reporting Resources

April S. Grady and Ronald S. Headrick

Introduction

A Brief Overview of Online Reporting Resources

Guidelines for Results-Oriented Online Documents

Guidelines for Interactive Results-Oriented Tools

Guidelines for Programmatic/Informational Web Pages

Guidelines for Site- and Score-Reporting Website Design

Developing Score-Reporting Resources: The Hurdles and Goals

Conclusion

Introduction

Score reporting is among the most challenging aspects of test development being testing agencies today. In order to be as helpful as possible to test stakeholders, test agencies must provide a valid and reliable measure of student performance, use a statistically sound method of performance with a single score or percentile, and provide the necessary information to help the scores be used in the proper context and to help the scores be used in the proper context.

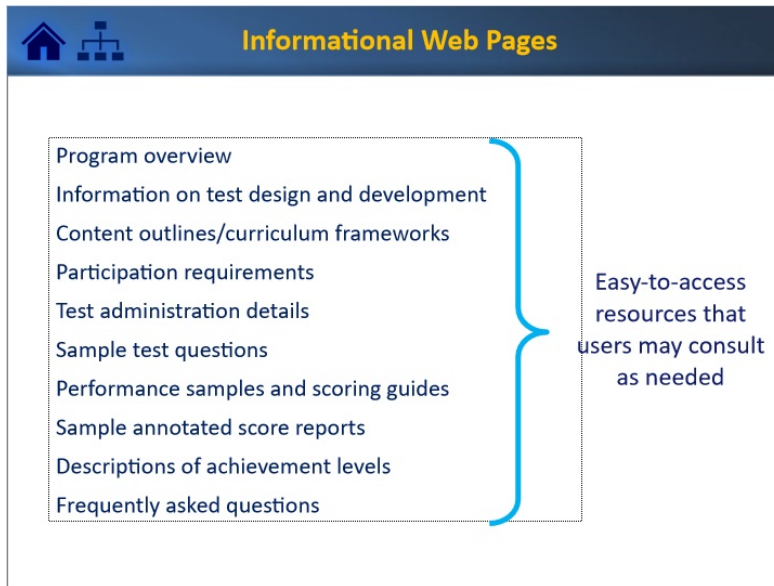
Content can mean many things, depending on the purpose of the test and the intended use of the data. Test results can be used to compare performance across groups, to identify areas of strength and weakness, and to provide a broad description of the distribution of scores. In addition, test results can be used to identify areas of strength and weakness, and to provide a broad description of the distribution of scores.

3.30 How IV

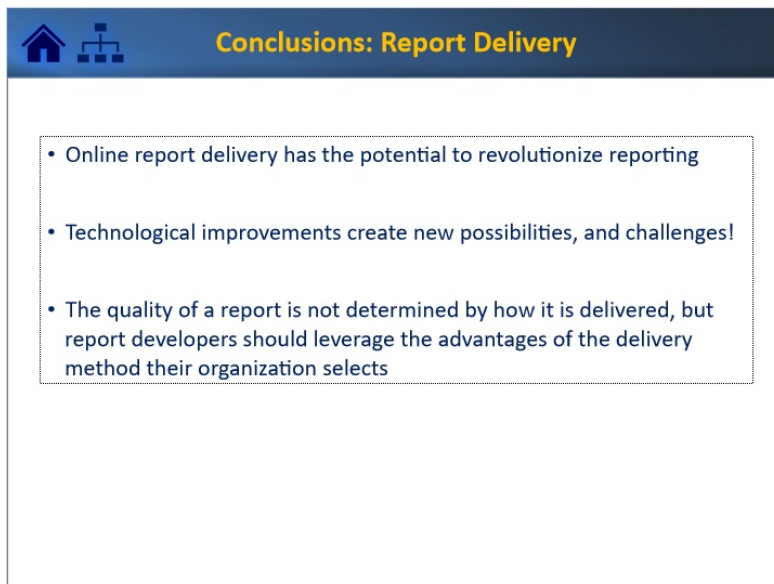
Possibilities for Online Reports and Tools

| | |
|---|--|
| <p>Greater customization</p> <p>On-demand interpretive information</p> <p>More convenience</p> <p>Increased user engagement</p> <p>New avenues for research</p> | <p>Confusion</p> <p>Difficulty accessing information</p> <p>Difficulty accessing reports</p> <p>Data overload</p> <p>Fairness issues</p> |
|---|--|

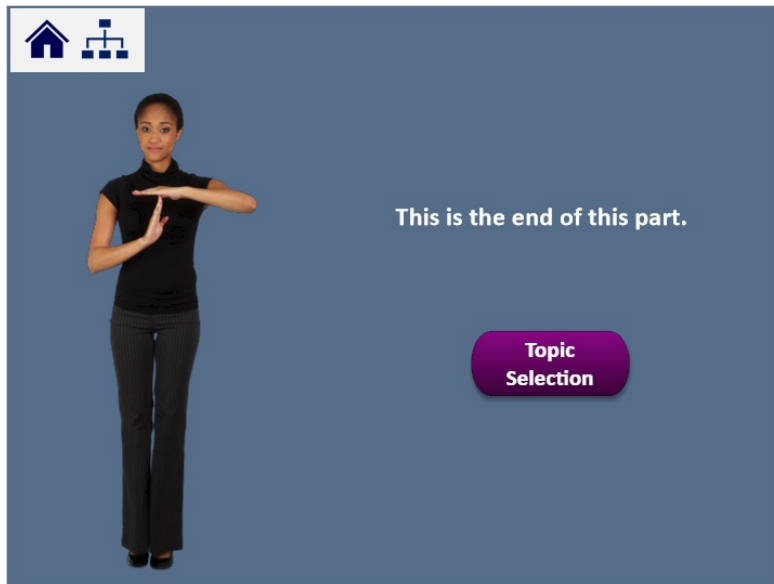
3.31 How V



3.32 How VI



3.33 Bookend: How





4. Section 3: Report Development


4.1 Cover: Section 3



4.2 Objectives: Section 3



Learning Objective



Develop a working understanding of how to apply an established model for the development, evaluation, and maintenance of results reports

4.3 Interview



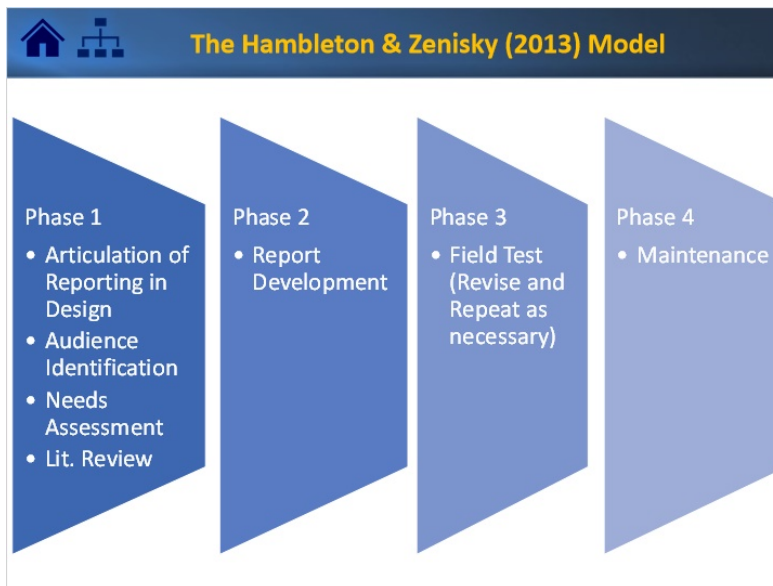
Section Structure: Interview with Dr. April Zenisky



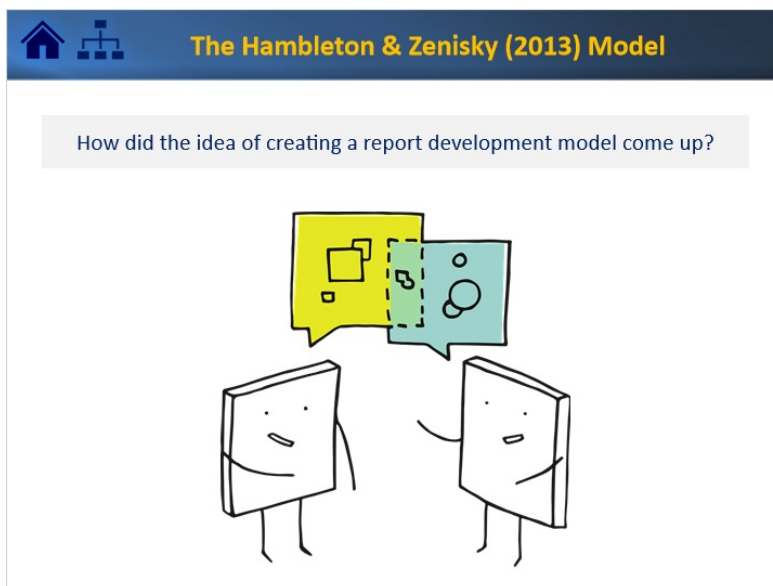
Dr. April Zenisky is one of the developers of the Hambleton and Zenisky (2013) model and a pioneer in the field of results reporting





4.4 Model



4.5 Model: How

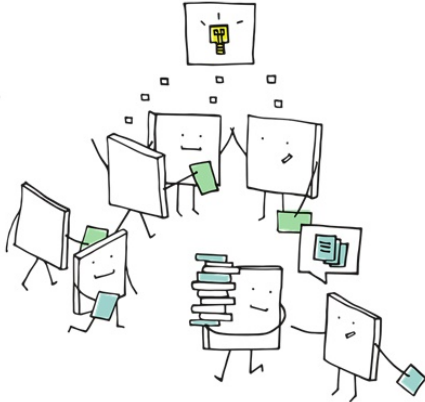


4.6 Model: Who





The Hambleton & Zenisky (2013) Model

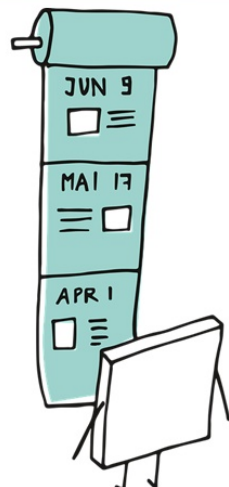
Who should use the Hambleton and Zenisky model?



4.7 Model: Timeline





The Hambleton & Zenisky (2013) Model



Is there a timeline attached to the model?

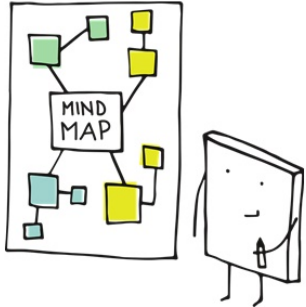
Do the steps have to be completed in order?

4.8 Model: Mix & Match





The Hambleton & Zenisky (2013) Model

A few other researchers have also proposed report development models.
Do report developers have to choose one? Is it okay to “mix and match”?




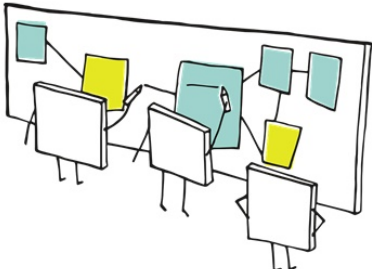
4.9 Model: Step 1



The Hambleton & Zenisky (2013) Model: Step 1

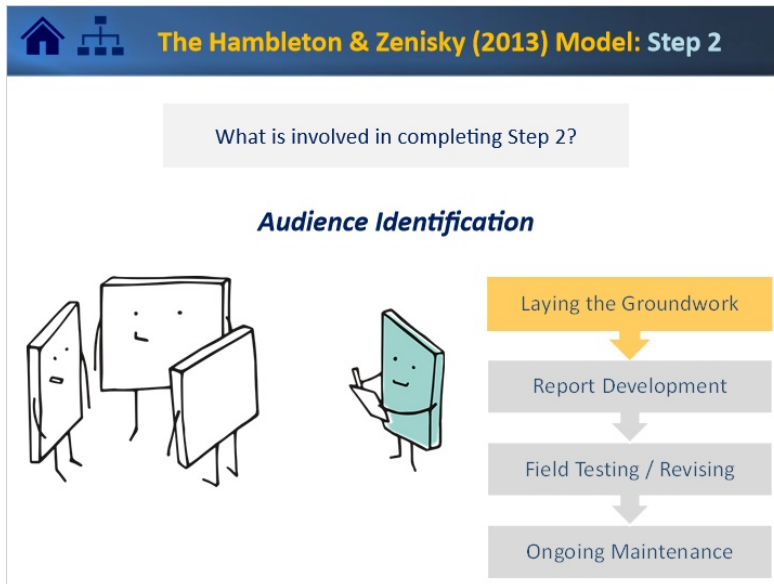
What are the core ideas of Step 1?

Articulation of Reporting in Design

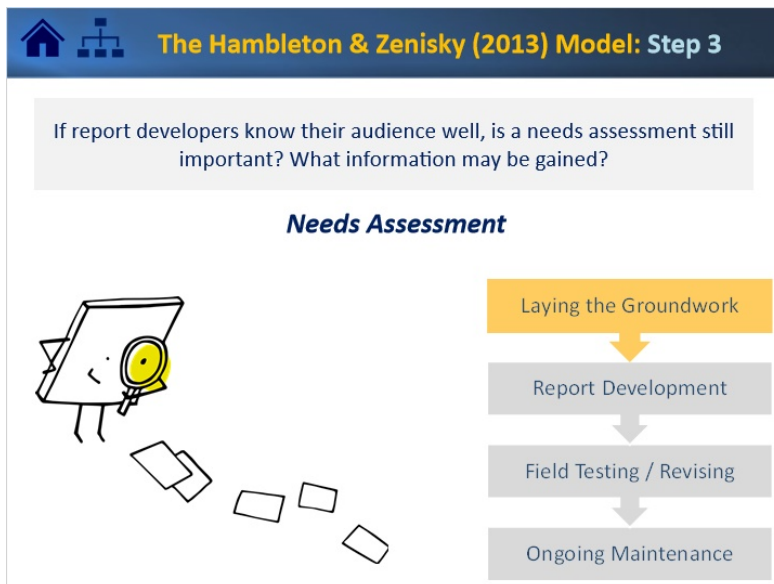


```
graph TD; A[Laying the Groundwork] --> B[Report Development]; B --> C[Field Testing / Revising]; C --> D[Ongoing Maintenance];
```

4.10 Model: Step 2



4.11 Model: Step 3




4.12 Model: Step 4

 **The Hambleton & Zenisky (2013) Model: Step 4**

Why is Step 4 important?
When do report developers know that they have enough information?

Literature Review





```
graph TD; A[Laying the Groundwork] --> B[Report Development]; B --> C[Field Testing / Revising]; C --> D[Ongoing Maintenance];
```

4.13 Model: Step 5

 **The Hambleton & Zenisky (2013) Model: Step 5**

Ideally, who should be involved in completing Step 5?
What are some pitfalls to avoid?

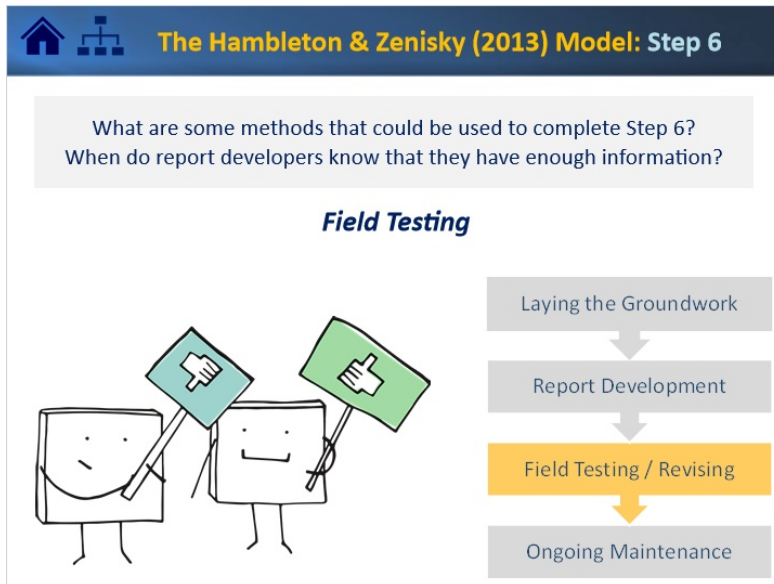
Report Development





```
graph TD; A[Laying the Groundwork] --> B[Report Development]; B --> C[Field Testing / Revising]; C --> D[Ongoing Maintenance];
```



4.14 Model: Step 6




4.15 Model: Step 7



4.16 Learning Check: Personal Reflection #4






Learning Check: Personal Reflection #4



- What phase of the Hambleton and Zenisky (2013) model are you most interested in or would you most like to be involved in? Why?
- What phase of the Hambleton and Zenisky (2013) model do you expect to be the most challenging for you or your organization? Why?

4.17 End of Section 3





This is the end of this section.

[Main Menu](#)

5. Section 4: Standards and Guidelines

5.1 Cover: Section 4



5.2 Objectives: Section 4

The image shows a slide titled "Learning Objectives" with a dark blue header. Below the header is a large white area containing a 3D target icon with an arrow hitting the bullseye. At the bottom of the slide, there are two blue-bordered boxes containing the learning objectives.

Learning Objectives

I. Understand major guidelines from professional organizations for developing and maintaining reports

II. Apply a working knowledge of empirical findings on report elements, layout, and appearance drawn from the psychometric literature to design results reports

5.3 Topic Selection

  **Introduction and Overview**





Guidelines
from
Professional
Organizations




Recommendations
from the
Research
Literature

Click on each light bulb to learn more 



5.4 Resource Overview (I)




Guidelines
from
Professional
Organizations

5.5 Resource Overview (I)



Key Resources

- ✓ **Standards for Educational and Psychological Testing**
(AERA, APA, & NCME, 2014)
- ✓ **Code of Fair Testing Practices in Education**
(Joint Committee on Testing Practices, 2004)
- ✓ **Standards for the Accreditation of Certification Programs**
(National Commission for Certifying Agencies, 2014)
- ✓ **Guidelines for Quality Control in Scoring, Test Analysis, and Reporting of Test Scores**
(International Test Commission, 2013)



5.6 Resource Selection



Key Resources



Standards



Code



Accreditation Standards



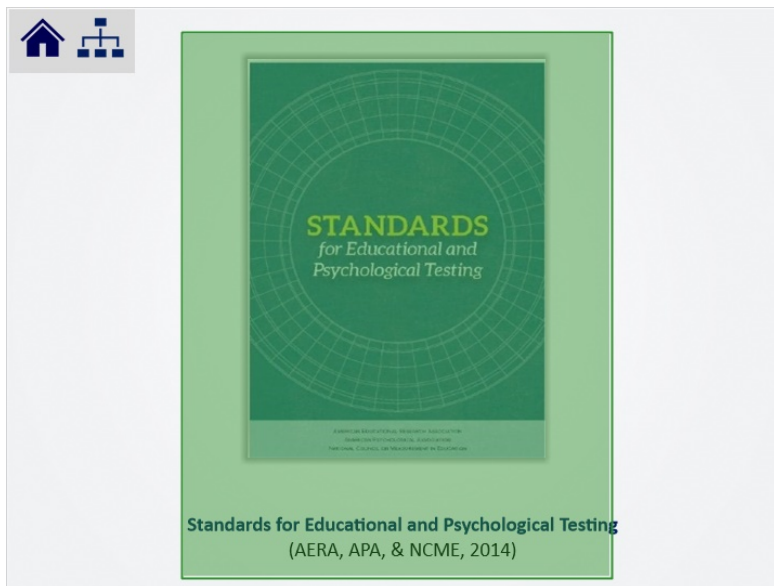
ITC Guidelines

Click on each image to learn more.




End Part



5.7 Bookmark: Joint Standards





5.8 AERA et al. (2014) Standards

**AERA et al. (2014) Standards**

- Offers detailed guidelines and commentary on the development and use of tests
- Prepared by the American Educational Research Association, American Psychological Association, and National Council on Measurement in Education with input from several organization
- Represents an authoritative source of collective knowledge and recommendations
- Available for purchase in English and Spanish



5.9 Standards Clusters





Cluster 1: Standards 6.10 - 6.16 (I)

Standards for Reporting and Interpretation

6.10: “When test score information is released, those responsible for testing programs should provide interpretations appropriate to the audience. The interpretations should describe in simple language what the test covers, what scores represent, the precision/reliability of the scores, and how scores are intended to be used.” (p. 119)

Commentary suggests that research be conducted to investigate if reports and other interpretive materials are interpreted as intended

5.10 Additional standards





Cluster 1: Standards 6.10 - 6.16 (II)

Additional standards for “Reporting and Interpretation” (Cluster 1):

- **6.11:** Ensuring the quality (especially validity) of automatically generated score interpretations
- **6.12:** Not reporting individual scores unless there is appropriate evidence to support individual-level interpretations of results
- **6.13:** Handling material errors in results reports
- **6.14:** Maintaining and using individually identifiable test score information
- **6.15:** Maintaining more detailed records if needed
- **6.16:** Protecting the confidential nature of test takers’ scores and ancillary information

5.11 Cluster 2: Standards 8.7-8.8



Cluster 2: Standards 8.7-8.8

Standards for “Test Takers’ Rights to Fair and Accurate Score Reports”

- **8.7:** Selecting appropriate labels and describing them precisely if scores are assigned to categories (i.e., achievement levels)
 - Appropriate labels are those that “reflect intended inferences” without being “unnecessarily stigmatizing” (p. 136)
- **8.8:** Providing test takers with timely access to reports, unless test takers have waived their right to this information explicitly or implicitly
 - Explicit versus implicit waiver
 - What is considered “timely access”?

5.12 Standards Clusters





Additional Standards



Two clusters of standards.... That’s all?

- **No.** All aspects of test development directly or indirectly influence, what information can be reported
- **Several standards in other clusters** provide useful recommendations for report development


5.13 Standard 1.1



Report Development: Standard 1.1

Consider Step 1 of the H&Z report development model and Standard 1.1:

- **Step 1:** Define purpose of report, connect to test purpose and interpretation/use argument.
- **Standard 1.1:** “The test developer should set forth clearly how test scores are intended to be interpreted and consequently used.” (p. 23)




5.14 AERA standards





Addressing Misinterpretations: Standard 5.3

Emphasis on safeguarding against possible misinterpretations and misuses of test results:




5.3: “If there is sound reason to believe that specific misinterpretations of a score scale are likely, test users should be explicitly cautioned.” (p. 102)

5.15 Recommendations to test users





Information Management

- The *Standards* makes several recommendations to provide certain types of information to test users
- They do NOT require that everything be included in the results report
- They emphasize that any information reported should be appropriate to its intended audience



5.16 AERA standard list

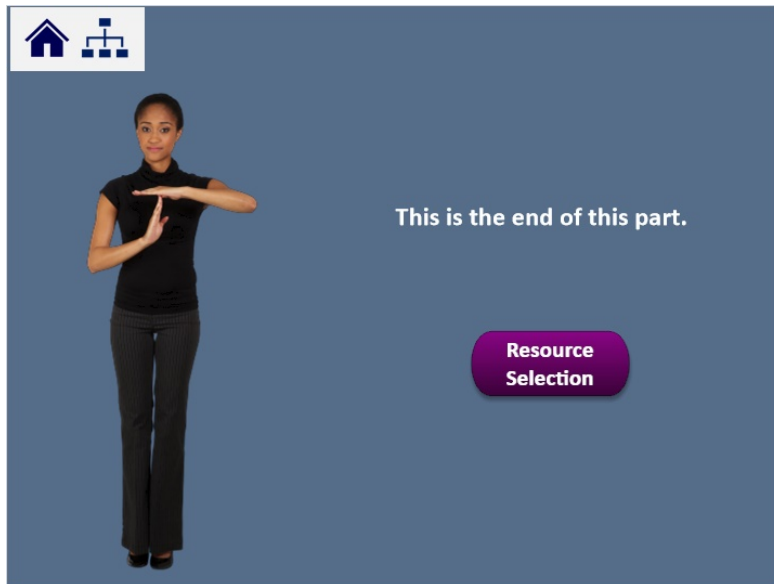


List of Applicable AERA et al. (2014) Standards

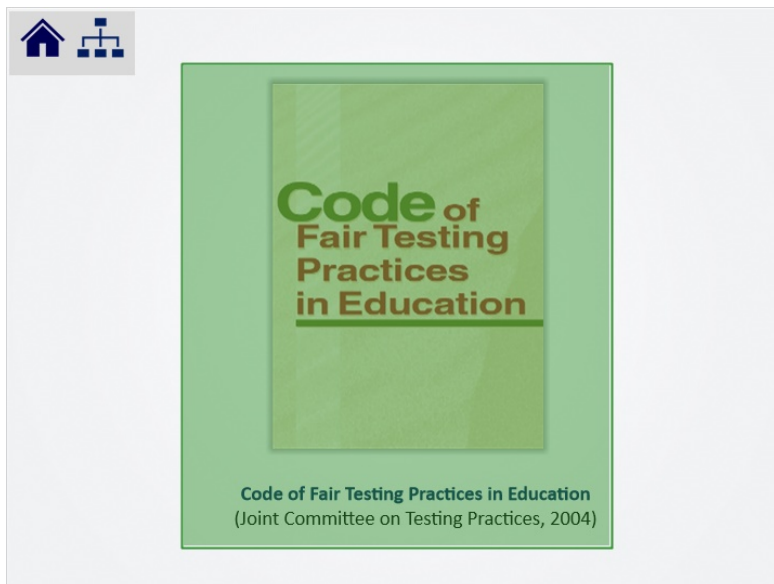
A review of all AERA et al. (2014) standards with recommendations that apply to results reporting is beyond the scope of this module but here is a list of relevant standards to consider for this work:

| | | | | |
|------|------|------|------|-------|
| 1.3 | 2.14 | 5.10 | 7.10 | 11.15 |
| 1.5 | 3.1 | 6.10 | 7.11 | 12.5 |
| 1.15 | 3.17 | 6.11 | 8.7 | 12.11 |
| 2.0 | 4.22 | 6.12 | 8.8 | 12.17 |
| 2.3 | 5.1 | 6.13 | 8.10 | 12.18 |
| 2.4 | 5.3 | 6.14 | 9.8 | 13.5 |
| 2.8 | 5.4 | 6.15 | 9.10 | 13.6 |
| 2.10 | 5.5 | 6.16 | 9.16 | 13.7 |
| 2.13 | 5.8 | 7.1 | 11.1 | 13.9 |



5.17 Bookend: Joint Standards



5.18 Bookmark: Code



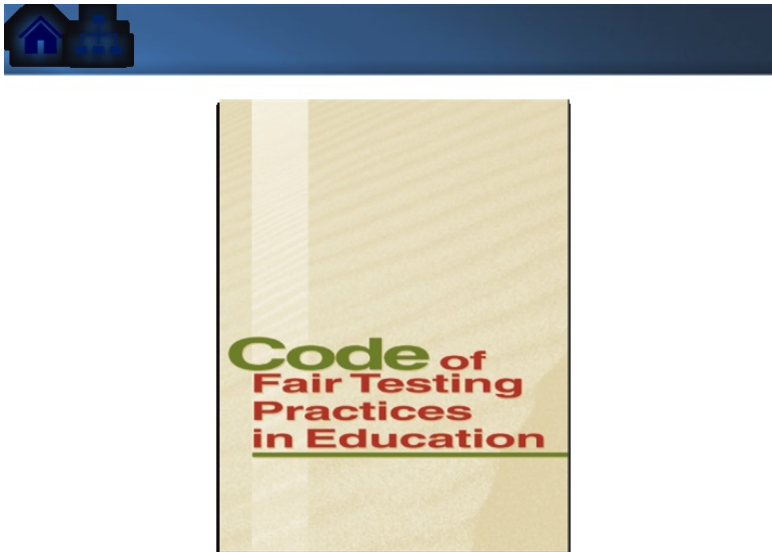
5.19 Code of fair testing




JCTP (2004) Code

- Prepared by the Joint Committee on Testing Practices
- Included the same organizations that worked on the AERA et al. *Standards* (2014) and others, such as the National Association of School Psychologists and the National Association of Test Directors
- A practical guide designed to cover testing in education but not employment, licensure, or certification testing
- Available as a free [PDF](#) on the website of the APA

pdf (Slide Layer)



5.20 Code Alignment




Alignment with AERA et al. (2014) Standards

Consistent with the 1999 *Standards*, but:

- much more condensed
- responsibilities shown separately for test developers and test users


Test developers:

“people and organizations that construct tests, as well as those that set policies for testing programs.” (p. 2)





Test users:

“people and agencies that select tests, administer tests, commission test development services, or make decisions on the basis of test scores.” (p. 2)



5.21 Code Organization



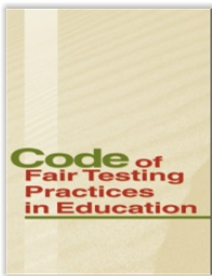
Organization: Four Topics

A. Developing and Selecting Appropriate Tests

B. Administering and Scoring Tests



C. Reporting and Interpreting Test Results

D. Informing Test Takers



- The last two areas are most applicable to results reporting
- We will focus on guidance for test developers

5.22 Reporting and Interpreting Results

 **Statements C1 & C3: Connections to Standard 6.10**

Topic C: “Reporting and Interpreting Test Results” (C1-C8)

AERA et al. Standard 6.10


When test score information is released, those responsible for testing programs **should provide interpretations appropriate to the audience**. The interpretations should describe in simple language **what the test covers, what scores represent**, the precision/reliability of the scores, and **how scores are intended to be used**.

Statement C1



Provide information to support recommended interpretations of the results, including the nature of the content, **norms or comparison groups**, and other **technical evidence**. **Advise test users of the benefits and limitations of test results and their interpretation**. Warn against assigning greater precision than is warranted.

Statement C3

Specify appropriate uses of test results and warn test users of potential misuses.





5.23 Other responsibilities

 **Reporting and Interpreting Test Results**

- **C2:** Provide guidance on interpreting results when test are administered with modifications
- **C4:** Present supporting evidence for any performance standards implemented and avoid the use of stigmatizing labels
- **C5:** Encourage test users to consider multiple sources of information to make decisions about a test taker
- **C6:** Provide information to support the interpretation of group-level results
- **C7:** Deliver results in a timely fashion and in a way that is understood by test takers
- **C8:** Offer guidance to test users on how to monitor the degree to which the test is fulfilling its intended purpose

5.24 Informing Test Takers





Topic D: Informing Test Takers

- No distinction between test user and test developer roles
- Introductory statement echoes many of the ideas in the AERA et al. (2014) *Standards*:

“Test developers or test users should inform test takers about the nature of the test, test taker rights and responsibilities, the appropriate use of scores, and procedures for resolving challenges to scores.” (p. 10)

5.25 Recommendations for test user and test developer





Informing Test Takers: Statements D1-D4

Recommendations cover test users’ or test developers’ responsibility to:


- **D1:** Inform test takers of key features of the test prior to test administration
- **D2:** Offer information to help test takers decide whether or not to take a test if it is optional
- **D3:** Communicate details about test takers’ rights pertaining to test administration (e.g., right to retake tests or have them rescored)
- **D4:** Inform test takers about their responsibilities (e.g., knowing the intended purposes and uses of the test, following directions, and performing at capacity)

5.26 Recommendation Continued




Informing Test Takers: Statements D5-D7

- **D5:** Let test takers know how and for how long results will be stored, and under what circumstances they may be released
- **D6:** Describe procedures for resolving issues that may result in canceling or withholding results
- **D7:** Provide details about how test takers may obtain more information about the test or register complaints



5.27 Bookend: Code






This is the end of this part.

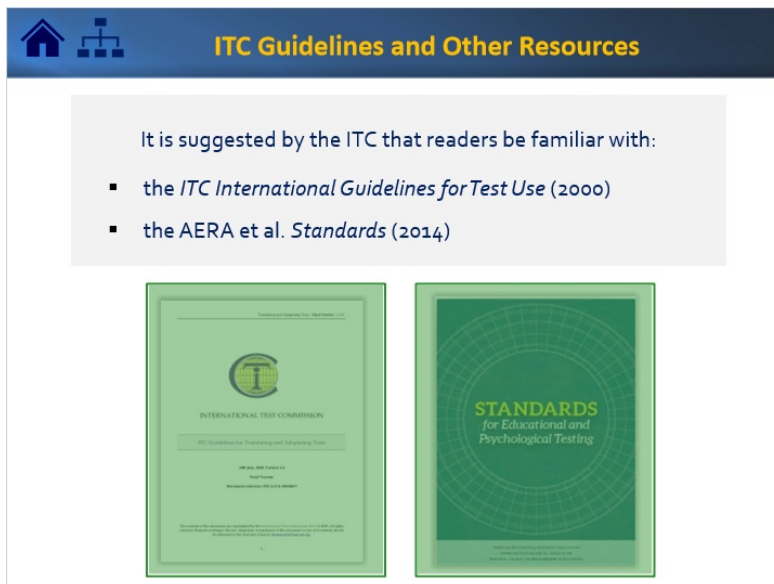
Resource Selection

5.28 Bookmark: ITC Guidelines





The screenshot shows the ITC Guidelines website. At the top left is a navigation icon. The main content area is titled "Guidelines" and lists various ITC guidelines. A sidebar on the left contains a list of guidelines: "The ITC Guidelines for Translating and Adapting Tests", "The ITC Guidelines on Test Use", "The ITC Guidelines on Computer-Based Testing", "The ITC Guidelines on Quality Control", "The ITC Guidelines on Test Security", "The ITC Guidelines on Test Disposal", "The ITC Guidelines for the Large-Scale Assessment of Diverse Populations", "The ITC Test-Taker's Guide", and "The ITC Statement on Using Tests for Research". The main content area is titled "Published Guidelines" and lists seven guidelines: 1. The ITC Guidelines for Translating and Adapting Tests, 2. The ITC Guidelines on Test Use, 3. The ITC Guidelines on Computer-Based and Internet-delivered Testing, 4. The ITC Guidelines on Quality Control in Scoring, Test Analysis and Reporting of Test Scores, 5. The ITC Guidelines on the Security of Tests, Examinations, and Other Assessments, 6. The ITC Guidelines on Practitioner Use of Test Revisions, Obsolete Tests, and Test Disposal, and 7. The ITC Guidelines for the Large-Scale Assessment of Linguistically and Culturally Diverse Populations. Below this list, it states: "You can download the Guidelines above, or visit their respective pages in this website." Further down, it mentions: "We are currently developing new guidelines on Technology-Based Assessments, in collaboration with the Association of Test Publishers (ATP)." At the bottom, it says: "Please cite the ITC guidelines as follows (example for Guidelines on Test Use): 'International Test Commission (ITC) (2000). International Guidelines for Test Use. Downloaded on [date] from www.intestcom.org'." The footer of the page reads: "Guidelines for Quality Control in Scoring, Test Analysis, and Reporting of Test Scores (International Test Commission, 2013)".

5.29 ITC Alignment



The screenshot shows the "ITC Guidelines and Other Resources" page. At the top left is a navigation icon. The main content area is titled "ITC Guidelines and Other Resources". Below this, it states: "It is suggested by the ITC that readers be familiar with:" followed by a list of two items: "the ITC International Guidelines for Test Use (2000)" and "the AERA et al. Standards (2014)". Below the list are two book covers. The left cover is the "ITC Guidelines for Translating and Adapting Tests" by the International Test Commission. The right cover is the "STANDARDS for Educational and Psychological Testing" by the American Educational Research Association (AERA) et al.

5.30 Clusters pertaining results reporting





ITC Guidelines for Reporting

Reporting test scores

2.5.1: “Ideally, score reports should be provided in printable form. In some instances and settings, the Internet is increasingly being used and is becoming the standard reporting **method**. **Reporting must be done in such a way that the meaning of the scores is clear to the test taker and to the client.**” (p. 214)

Printed score reports that arrive in the mail are becoming less common , but most reports still follow a printable format.

5.31 Example of methods





Development of Interpretive Guides

Data Collection Methods

2.5.1.1: “Use focus groups of test takers or possibly ‘think-aloud procedures,’ ‘experimental studies,’ or even ‘one-on-one interviews’ to gain information to assist in the development of comprehensible and instructive explanations of the score report and any interpretive guide.” (p. 214)

The same methods could be used to support the development and refinement of reports.

5.32 Guidelines of responsibilities





Responsibilities and What to Include

2.5.1.2: “. . . anyone who receives the scores has appropriate guidance in interpreting them,” and evidence should be collected to show that “reports allow users to make defensible interpretations” (p. 214)

2.5.1.3: “Create computer-generated reports with an explanation of the test and the meaning of scores that can help mediate technical issues and that will e appropriate for their recipients.” (p. 214)

5.33 ITC Guidelines recommendations





Additional Recommendations

The *ITC Guidelines* recommend clarifying appropriate uses for different types of scores, taking precision into consideration **(2.5.1.5)**

Other guidelines address:

- Using data warehouses as needed **(2.5.1.4)**
- The possibility of enlisting the help of public relations experts to communicate results to the media and policymakers **(2.5.1.6)**
- Several steps for maintaining the security of results reports **(2.5.2.1-2.5.2.5)**



5.34 Bringing it all together




Bringing it All Together: Conclusions

- The AERA et al. (2014) *Standards* should be “**required reading**”
- Other sources (*JCTP Code, NCCA Standards, and ITC Guidelines*) also provide valuable information and may be required reading **depending on one’s field**
- None of the sources provide step-by-step instructions for developing a perfect report (there is **no one-size-fits-all solution!**)
- All sources provide key recommendations for developing reports that **support appropriate interpretations and uses** and diminish the likelihood of misinterpretations and misuses
- We can turn to the **research literature** for additional recommendations

5.35 Learning check: Reflection

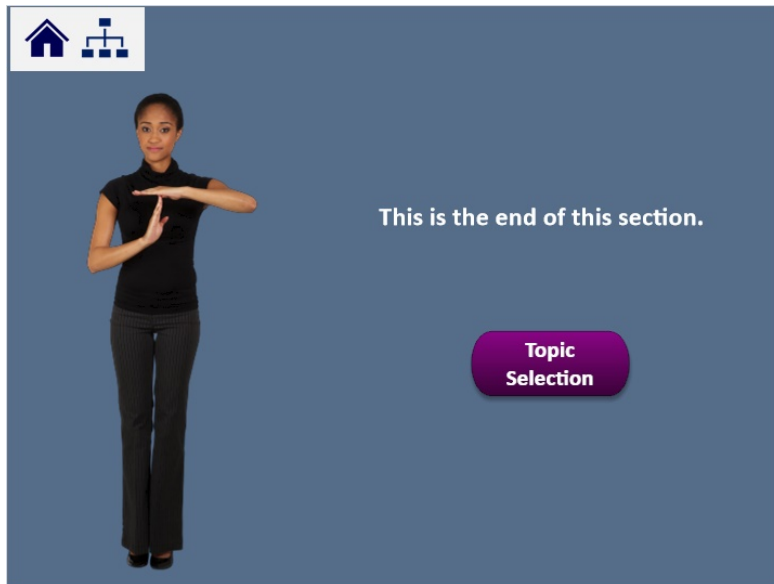


Learning Check: Personal Reflection #5



In reflecting on the various standards presented, how do you see yourself using each of the resources - or not using them - in your reporting efforts in the future?

5.36 Bookend: ITC Guidelines



5.37 Bookmark: Accreditation Standards



5.38 Standards for the Accreditation of Certification Programs (2014)





NCCA (2014) Accreditation Standards


- Prepared by the National Commission for Certifying Agencies, the accreditation body of the Institute for Credentialing Excellence (ICE)
- Created in the mid-1970s and updated periodically
- Emphasize the essential features of a high-quality certification program
- Free to members of ICE, available to others for a fee



website (Slide Layer)



Standards for the Accreditation of Certification

 Institute for Credentialing Excellence

Education Accreditation Resources Career Center Community Research ICE Exchange Store

Twitter Facebook LinkedIn YouTube


Promoting Best Practices for the Credentialing Community

The Institute for Credentialing Excellence provides education, networking, and other resources for professionals.

[About ICE](#) [Join ICE](#) [Directories](#)


[Join Now](#)

5.39 ITC Guidelines





ITC Guidelines (2013)

- Prepared by the International Test Commission (ITC)
- Quality-control guidelines to improve the efficiency, precision, and accuracy of testing processes, including results reporting
- Available for free on the ITC website




Pdf (Slide Layer)



ITC Guidelines

Quality Control in Scoring, Test Analysis, and Reporting of Test Scores - Final Version - v 1.2





INTERNATIONAL TEST COMMISSION

ITC Guidelines on Quality Control in Scoring, Test Analysis, and Reporting of Test Scores

12th October, 2013, Version 1.2
Final Version
Document reference: ITC-G-QC-20130102

The contents of this document are copyrighted by the International Test Commission (ITC) © 2013. All rights reserved. No part of this document may be reproduced or transmitted in any form or by any means electronic or mechanical, including photocopying and recording, or by any information storage or retrieval system, without prior written permission from the ITC.

5.40 Alignment NCCA





Connections

- The *NCCA Standards* (2014) are consistent with the AERA et al. *Standards* (2014), but focus on the certification context
- Like the AERA et al. *Standards*, the *NCCA Standards* offers guidelines (“essential elements”) along with commentary



5.41 NCCA Standards





Essential Elements D & E (Standard 19)

In certification, differential reporting based on whether examinees passed or did not pass is a sensible possibility

D: “All candidates must be provided information on their overall performance on an examination.” (p. 26)

E: “Failing candidates must be provided with information about their performance in relation to the passing standard. If the program provides feedback to candidates such as domain-level information, candidates must be provided guidance about limitations in interpreting and using that feedback.” (p. 26)

5.42 NCCA Standards 19



Commentary on Reporting Responsibilities

“The certification program should provide candidates with an explanation of the types of scores reported, appropriate uses, and potential misuses”
(p. 26)

“If domain-level information has low reliability, programs are advised against reporting it to candidates and other stakeholders. When domain-level or other specific feedback is given to candidates, the certification program should provide estimates of its precision and/or other guidance.”
(p. 26)

5.43 NCCA standards II

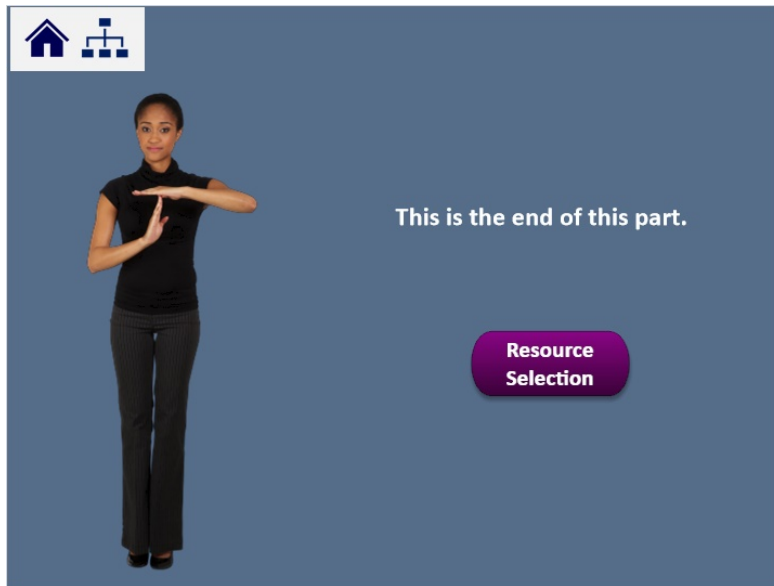


Procedural Aspects of Reporting (Standards 9-10)

- **Standard 9:** How long to maintain examinees’ test results and other certification-related records
- **Standard 10:** How to adequately protect the confidentiality of examinees’ results and related information



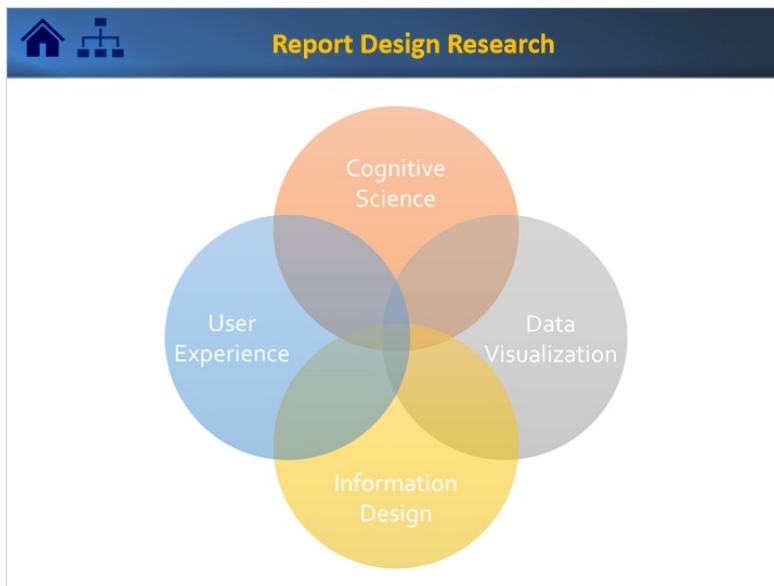
5.44 Bookend: Accreditation Standards



5.45 Bookmark: Best Practices for Reporting Results



5.46 Report Design Research

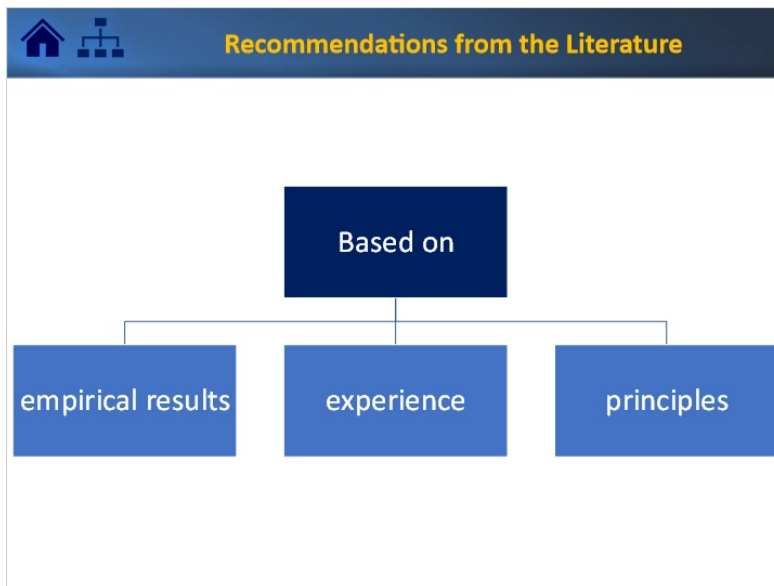


5.47 Generalizability

There isn't **one** perfect spaghetti sauce...
...or results report!

Previous studies offer useful ideas and recommendations, but be mindful of your assessment specifics and report users' needs



5.48 Recommendations from Literature




5.49 Topic Selection



5.50 Introductory Materials



Introductory Materials





Provide sufficient information about the purpose of the test
(Zenisky & Hambleton, 2012)

Use a simple narrative to summarize key results
(Zenisky & Hambleton, 2012)

[Back to Recommendations](#)

5.51 Overall Results I



Overall Results



Include complete information about the meaning and limitations of each score, but no more information than needed by report users
(Hattie, 2009; Hegarty, 2018)

Keep statistical jargon to a minimum
(Hambleton & Zenisky, 2013; Zwick, Zapata-Rivera, & Hegarty, 2014)

Tailor the complexity of statistical information to report users
(Hullman, Rhodes, Rodriguez, & Shah, 2011)

If misinterpretations are likely, use text interpretations to replace or support numerical results
(Goodman & Hambleton, 2004; Hattie, 2009)

5.52 Overall Results: Graphics




Overall Results: Graphics

Ensure that text and graphic elements are sufficiently large and consistent with data visualization principles
(Hegarty, 2018; Hullman et al., 2011; Slater, Livingston, & Silver, 2018)

Use graphics that are clear, simple, and satisfy specific purposes
(Goodman & Hambleton, 2004; Ryan, 2006; Wainer, 1997)

Do not rely too heavily on legends and interpretive guides
(Hegarty, 2018)

Use text to support the interpretation of graphics
(Goodman & Hambleton, 2004; Ryan, 2006)



5.53 Overall Results: Caution



Overall Results: Precision



Be aware that report users prefer different representations of error depending on their level of comfort with statistics
(Zwick et al., 2014)

The question of whether information about measurement error is beneficial to report users has no definitive answer
(Kannan, Bryant, Zapata-Rivera, & Peters, 2017; Kannan, Zapata-Rivera, & Leibowitz, 2018; Zapata-Rivera, Kannan, & Zwick, 2018)



Back to Recommendations

5.54 Performance Level



Performance Levels

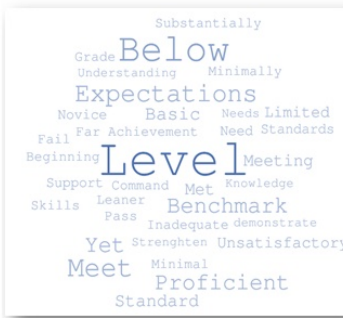
Recommendations for selecting performance level labels from a study that included teachers (N = 51), parents (N = 50), and students (N = 24):

Labels **should not**:

- Sound like character insults
- Be overly positive
- Diminish students' role in the learning process

Labels **should**:



- Convey a sensible amount of positivity or have a neutral tone
- Be consistent across grades
- Be consistently clear across levels of performance



Substantially
Grade Below
Understanding Minimally
Expectations
Novice Basic Needs Limited
Fail Far Achievement Need Standards
Beginning Level Meeting
Support Command Met Knowledge
Skills Learner Benchmark
Pass Inadequate demonstrate
Yet Strengthen Unsatisfactory
Meet Minimal
Proficient
Standard

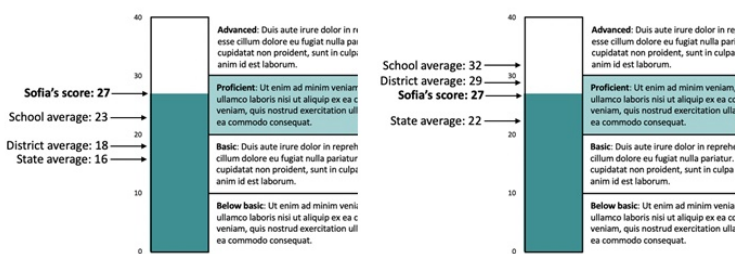
[Back to Recommendations](#)

5.55 Reference group comparison information



Reference Group Comparison Information

The prominence of reference group comparison information should **match the purpose of the assessment**



Left Chart: Sofia's score: 27. School average: 23. District average: 18. State average: 16. The bar chart shows Sofia's score is significantly above the school, district, and state averages.

Right Chart: School average: 32. District average: 29. Sofia's score: 27. State average: 22. The bar chart shows Sofia's score is below the school and district averages but above the state average.

Advanced: Duis aute inure dolor in n esse cillum dolore eu fugiat nulla pari cupidatat non proident, sunt in culpa anim id est laborum.



Proficient: Ut enim ad minim veniam, ullamco laboris nisi ut aliquip ex ea c veniam, quis nostrud exercitation ull ea commodo consequat.

Basic: Duis aute inure dolor in repreh cillum dolore eu fugiat nulla pariatur cupidatat non proident, sunt in culpa anim id est laborum.

Below basic: Ut enim ad minim veni ullamco laboris nisi ut aliquip ex ea c veniam, quis nostrud exercitation ull ea commodo consequat.


[Back to Recommendations](#)

5.56 Content area results





Content Area Results

- Use different scales for the total score and content area scores
(Haberman & Sinharay, 2010)
- Provide sufficient information about error
(Goodman & Hambleton, 2004; Zenisky & Hambleton, 2012)
- Indicate the number of items comprising each content area
(Goodman & Hambleton, 2004)
- Ensure that what is considered weak or strong performance is consistent
(Sinharay, Puhon, Haberman, & Hambleton 2018)




5.57 Content area results II





Content Area Results

- Use analogies to facilitate comprehension of complex content area results
(Sinharay, Puhon, Haberman, & Hambleton, 2018)
- Provide interpretive information and a visual summary of performance
(Roduta Roberts & Gierl, 2010)
- Include a summary of how any diagnostic content area results should be used
(Roduta Roberts & Gierl, 2010)




Back to Recommendations

5.58 Growth results



Growth Results

- Use analogy to make the concept of growth more accessible (Zenisky & Hambleton, 2014)
- Include a clear description of who is in the norm group for SGPs (Zenisky, Keller, & Park, 2018)
- Use line graphs to facilitate comparisons of growth over time (Hegarty, 2018; Zenisky et al., 2018)
- If information is provided in text, ensure that it is centrally located and that the font size is sufficiently large (Zenisky et al., 2018)
- Provide information about measurement error (Zenisky & Hambleton, 2014)



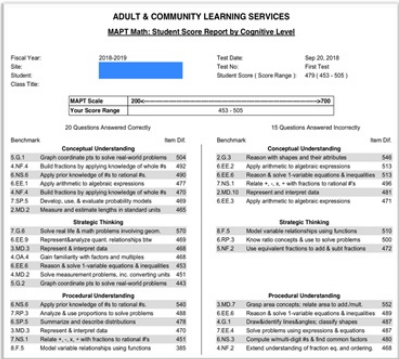
[Back to Recommendations](#)

5.59 Individual test question results

Individual Test Question Results

Use meaningful categorizations such as correct/incorrect by content area or Wright and Stone's (1979) four categories (see also: Brown, O'Leary, & Hattie, 2018)



correct, hard

incorrect, hard

correct, easy

incorrect, easy

[Back to Recommendations](#)

5.60 Organizational devices



Organizational Devices (I)





Make the report concise and uncluttered
(Goodman & Hambleton, 2004; Hattie, 2009; Slater et al., 2018)

Present information in a logical order
(Goodman & Hambleton, 2004; Hattie, 2009)

Use headings, boxes, shaded areas, and/or white space to organize the report
(Goodman & Hambleton, 2004; Slater et al., 2018)

5.61 Organizational devices II




Organizational Devices (II)



Favor bullet points over long sentences
(Rick, Slater, Kannan, Sireci, Zenisky, & Dickey, 2016)

Use embellishments such as icons only if they improve comprehension
(Slater et al., 2018)

Ensure that colors have adequate contrast, are used consistently, convert well to B&W, and do not limit readability for individuals with color blindness
(Trout & Hyde, 2006)



5.62 Learning Check: Personal Reflection #6



Learning Check: Personal Reflection #6

Consider the 8 areas which were discussed in this module.

1. Introductory materials
2. Overall results
3. Performance levels
4. Reference group comparison information
5. Content area results
6. Individual test question results
7. Growth results
8. Organizational devices

In thinking about a results report you work on or are familiar with, jot down one or more key actions you could consider in each area (where relevant) to improve the report.

[Back to Recommendations](#)

6. Section 5: Research on Reporting

6.1 Cover: Section 5







Section 5:


Conceptualizing Research on Results Reporting

[20 Minutes]

6.2 Objectives: Section 5





Learning Objectives



- I. Identify the research traditions reflected in studies on results reporting
- II. Articulate the core ideas of three conceptual frameworks
- III. Develop a conceptual foundation for planning and executing research on results reports

6.3 Research Questions and Methods





Research Questions and Methods

We choose research methods that are best suited to address our questions, so where do questions about results reporting come from?

They tend to be influenced by areas for improvement that psychometricians encounter in practice, such as:

- Validity
- Reliability
- Score comparability
- Content area results

6.4 Areas of focus





Area of Focus

Most publications on results reporting focus on **how to best communicate the assessment concepts we know to lay audiences**

Developing test **score reports** that work: The process and best practices for effective communication
Al Zensky, RK Hambleton - Educational Measurement: Issues ..., 2012 - Wiley Online Library
Test scores matter these days. Test-takers want to understand how they performed, and test **score reports**, particularly those for individual examinees, are the vehicles by which most people get the bulk of this information. Historically, **score reports** have not always met the ...
☆ 99 Cited by 70 Related articles All 3 versions
Comparing graphical and verbal representations of measurement error in test **score reports**
R Zwick, D Zapata-Rivera, M Hegarty - Educational Assessment, 2014 - Taylor & Francis
Research has shown that many educators do not understand the terminology or displays used in test **score reports** and that measurement error is a particularly challenging concept. We investigated graphical and verbal methods of representing measurement error ...
☆ 99 Cited by 33 Related articles All 3 versions
Keeping your audience in mind: Applying audience analysis to the design of interactive **score reports**
JD Zapata-Rivera, JR Katz - Assessment in Education: Principles ..., 2014 - Taylor & Francis
Score reports have one or more intended audiences: the people who use the reports to make decisions about test takers, including teachers, administrators, parents and test takers. Attention to audience when designing a score report supports assessment validity by ...
☆ 99 Cited by 37 Related articles All 3 versions
Does interpretive test score information help teachers?
JC Impara, KP Divine, FA Bruce, ... - Issues and Practice, 1991 - psycnet.apa.org
Assessed the adequacy of interpretive information (NI) provided on the reverse side of the Virginia State Assessment Program (VSAP) student **score reports**. 279 teachers completed a 17-item questionnaire and a score report for a hypothetical student ...
☆ 99 Cited by 48 Related articles All 2 versions

A few results from a Google Scholar search

6.5 Areas of focus



A Review of Research from 2005 to 2015



Gotch and Roduta Roberts (2018) reviewed papers on individual-level results reports published or presented between 2005 and 2015

60 papers out of an initial pool of 358 were analyzed, all of which:

1. Described an investigation of some aspect of results reporting
2. Focused on a formal reporting of individual-level results

Papers were coded based on several characteristics, including their stated or implied **theoretical framework of communication**

6.6 Findings



Findings from Gotch and Roduta Roberts (2018)

58 out of 60 papers aligned with the cybernetics tradition

- Investigations of what information to present and how to present it to maximize accurate interpretations



1 aligned with the semiotics tradition

- An investigation of what is implicitly communicated by elements such as space allotment and word choice

1 aligned with the sociopsychological tradition


- An investigation of how different types of feedback influence students' use of cognitive regulation strategies

6.7 Communication theory traditions (part 1)




Communication Theory Traditions (Part 1)


Cybernetics: Communication as information processing



Semiotics: Communication as a process that relies on signs





Rhetoric: Communication as artful, persuasive discourse



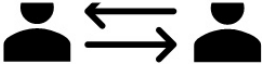
Craig (2007, 2009)

6.8 Communication theory traditions (part 2)




Communication Theory Traditions (Part 2)



Sociopsychological: Communication as social interaction and social influence



Sociocultural: Communication as an interactive process that creates and reinforces social reality, including rituals and structures




6.9 Communication theory traditions (part 3)






Communication Theory Traditions (Part 3)



Phenomenology: Communication as the experience of self and other in dialogue



Critical: Communication in which assumptions can be freely interrogated to promote social justice and mutual understanding





6.10 Conceptual layers of reporting



Sample Questions

| | |
|--|---|
| Semiotics Are certain color schemes for different levels of performance more encouraging than others? | Rhetoric How do different introductory messages in reports affect users' trust in the results? |
| Sociopsychological How does the influence of results reports compare to the influence of written feedback from teachers? | Sociocultural Are students more likely to reflect on feedback from results reports if teachers model that behavior? |
| Phenomenology How do students experience the day when they receive results reports from high-stakes tests? | Critical Are there ways of reporting assessment results that better support social justice? |

6.11 Conceptual Layers of Reporting




Conceptual Layers of Reporting

Behrens, DiCerbo, Murphy, and Robinson (2013) proposed **three conceptual layers** for understanding the relation between assessment developers and assessment stakeholders.

1. Information communication


2. Social activity

3. Educational literacy / Societal transformation

Click on each layer above to view 

Information communication (Slide Layer)

Information Communication



Core question:
"What do designers want to communicate?"

Interpretive constraints

- Establish what interpretations can be supported by data

Communicative considerations


- Design reports that promote those interpretations, following design recommendations and considering report users' typical literacy and numeracy

Process imperatives

- Gather evidence of how the reports function

Educational literacy / Societal transformation (Slide Layer)

Educational literacy / Societal transformation



Core question:
"What is the desired effect in the educational system?"

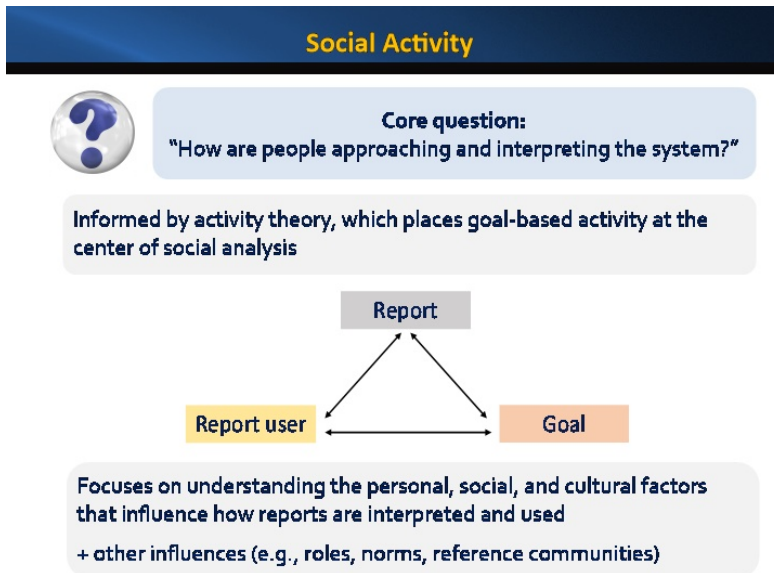
Educational literacy: "The large scale understanding of how learning outcomes are shaped and changed by practices." (p. 18)

Emphasizes the need to articulate the short- and long-term goals of using a particular assessment, at the micro and macro levels

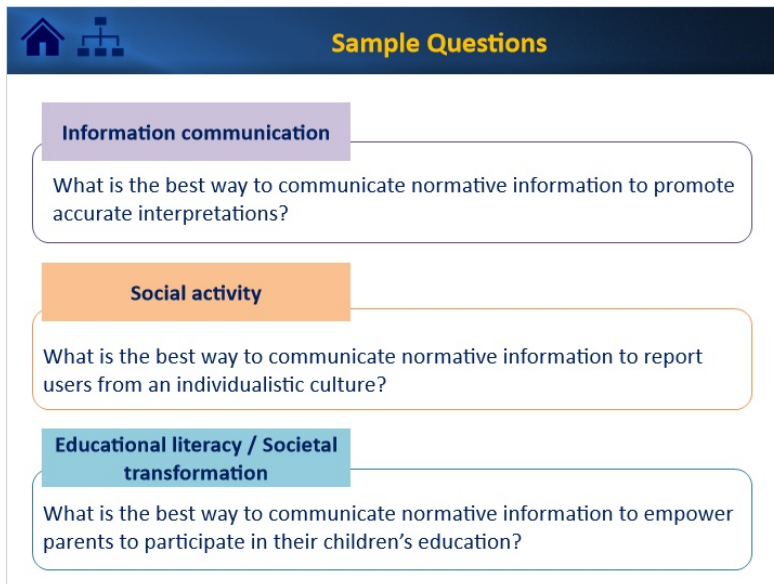
Raises questions about:

- How results reports and related assessment information influence the education system
- How knowledge from assessments interacts with information about instructional standards and practices

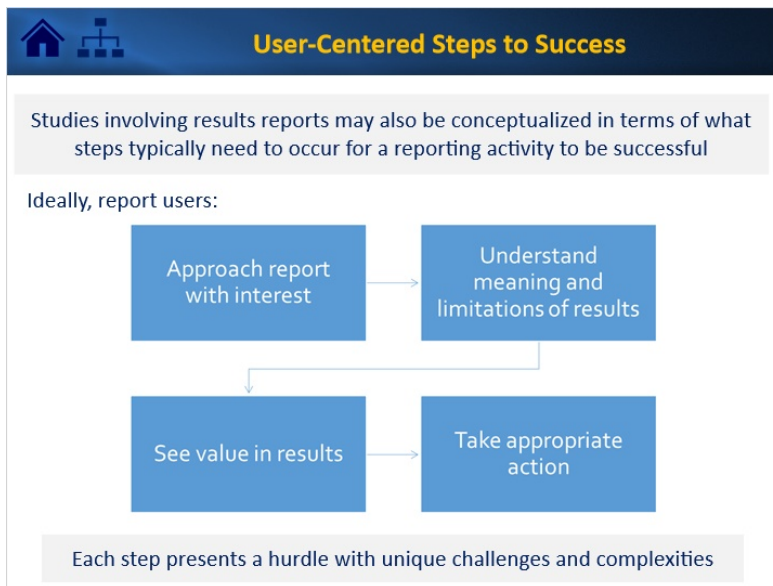
Social activity (Slide Layer)



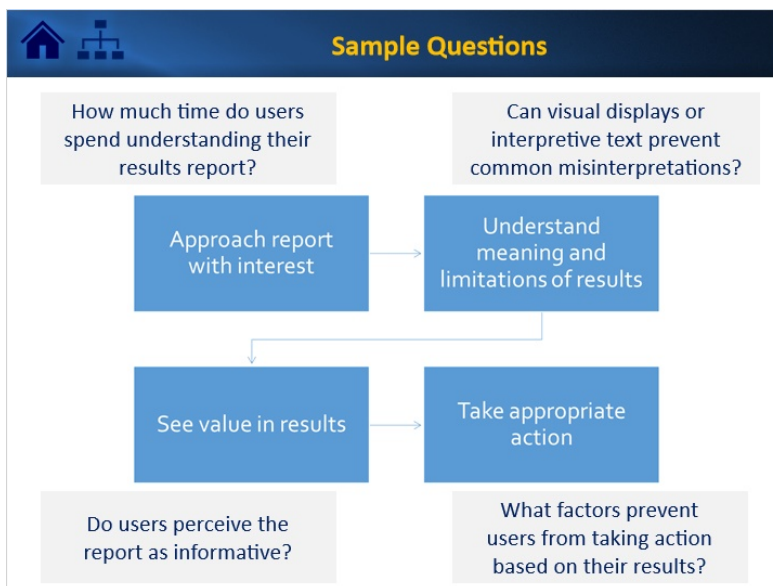
6.12 Sample Questions



6.13 User-Centered Steps to Success



6.14 Sample Questions



6.15 Sample Questions



Conclusion



Research on results reporting can evolve in multiple directions

For conceptual frameworks and ideas, consider:


- Applications of communication theory traditions to results reporting
 - Gotch and Roduta Robert's (2018) review is a valuable starting point
- Behrens, DiCerbo, Murphy, and Robinson's (2013) Conceptual Layers of Reporting
- User-Centered Steps to Success



6.16 Personal Reflection #7

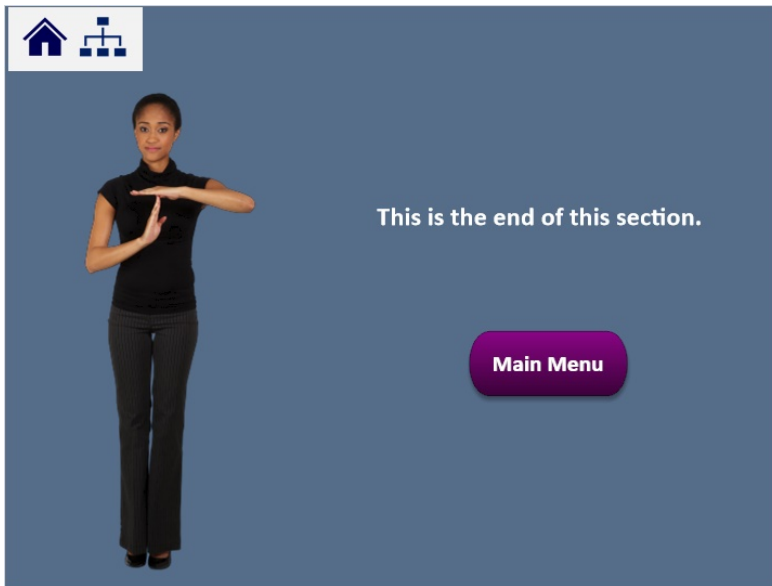


Learning Check: Personal Reflection #7



Choose a conceptual framework listed on slide 6.17, and briefly describe the major points of that framework as they relate to research on reporting that you might do or be interested in.

6.17 End of Section 5



6.18 Module Cover (END)

