Vertically-Articulated Content Standards

Lauress Wise

Human Resources Research Organization (HumRRO)

Edward F. Reidy, Jr. Interactive Lecture Series

Nashua, New Hampshire

October 7, 2004

Vertically-Articulated Content Standards

Presentation outline

- ➢ **So you want to measure student growth?**

- ➢ **You should know about CCSSO/TILSA work on vertical alignment!**

- ➢ **Let's do a sample exercise.**

- ➢ **Further work is needed.**

Measuring Student Growth

- **Current pressure to build vertical scales!**
  - Report "growth" for individual students.
  - Assess each school's "value added"
    - ✓ Gains versus status
    - ✓ Are students learning (not have they learned)
- **But what does the vertical scale measure?**
  - Combined curriculum content across grades?
  - How can scale points be described in terms of content standards for different grades?

What Is Vertical Alignment?

Vertical alignment asks:

➢ How are content standards/objectives related from one grade to the next?

- ■ Knowledge or skills extended to **wider range** of content

- ■ **Deeper** understanding (cognitive processes) for the same content

- ■ **New** content and/or skills

TILSA Work on Vertical Alignment

➤ Initial focus on supporting vertical scales

  ■ Is content alignment sufficient to justify a vertical scale?

  ■ How to label points along the vertical scale?

➤ Changed to focus on quality of vertical articulation

  ■ Concerns about misuse of vertical scales

    ✓ Inferences about mastery of content not tested

    ✓ Scales will vary by content of items used in linking

  ■ Other important needs for clarifying content standards and their relationship across grades

    ✓ Helping teachers talk "across grades"

    ✓ Clarifying test specifications within each grade

    ✓ Supporting the development of curriculum materials

Nature of Content Alignment

Applying Webb's Alignment constructs

1. **Categorical Concurrence**

   - What content is new? What content is continued?

2. **Range of Content**

   - Broadening or generalizing knowledge/skills

3. **Depth of Knowledge (DOK)**

   - Webb DOK ratings are somewhat grade-specific.

4. **Balance of Representation**

   - How does content emphasis vary across grades?

5. **Source of Challenge**

   - What needs to be clarified about the standards?

Quality of Content Alignment

➤ Content standards are ***not*** clearly articulated across grades if:

- Related standards are not clearly differentiated.
  - ✓ What new knowledge or skill is required?
  - ✓ One or both standards may not be described in sufficient detailed.
- Differences in terminology are not explained.
  - ✓ Different words for the same skill?
- Terminology drifts.
  - ✓ The meaning of terms appears to be expanded.
- Specific objectives are omitted at some grade.

Gathering Content Alignment Data

- ➤ Who should judge?
  - ▪ Same experts who developed the content frameworks.

- ➤ What are judges asked to do?
  - ▪ Make judgments about individual standards.
    - ✓ Grade-to-grade comparisons (summed up later)
    - ✓ Within specific content areas or subscales
      - • To limit search for similar standards
  - ▪ Identify related prior-grade standard(s)
  - ▪ Describe relationship
    - ✓ Qualitative description of what is new or added.
    - ✓ Code relationship type (Extend, Deeper, New, Same, Prerequisite)
  - ▪ Identify quality issues
    - ✓ Source(s) of challenge

Reporting Vertical Alignment

- ➢ Detailed reports
  - ▪ Content Maps
  - ▪ List of specific challenges (articulation quality concerns)
- ➢ Summary indicators
  - ▪ Concurrence - % new content
  - ▪ Range - % of skills broadened
  - ▪ Depth - % of skills deepened
  - ▪ Balance - % of standards with few/many objectives
 - ▪ Challenge – Average rating; % flagged with comments

Simplified Content Map

Grade 1	Grade 2	Grade 3
1.1 Add and subtract integers	2.1 Add and subtract decimals. (E)	3.1 Apply arithmetic operations to solve problems (D)
1.2 Multiply and divide integers	2.2 Multiply and divide decimals (E)	3.2 Apply arithmetic operations to fractions (E, D)
	2.3. Recognize negative numbers. (N)	3.3 Define rational and irrational numbers (E)
		3.4 Understand basic rules of exponents (N)

Sample Exercise

➤ To be distributed at the conference (time permitting).

Next Steps

➢ Complete concept paper for the current project.

➢ Identify opportunities for further pilot work.

- ▪ Improve data collection protocols.

- ▪ Develop/improve rater training.

- ▪ Build detailed examples of reports.

- ▪ Begin to talk about more specific standards for good vertical alignment.
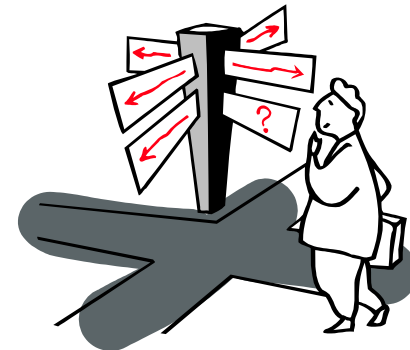
Checking Your Vertical Alignment

> **Do your standards need an alignment checkup?**

- Identify needs for revision.
- Add explanatory text.
- Define common or evolving content for a vertical scale.
- Suggest labels for points on the vertical scale.
- Satisfy NCLB requirements for coherent grade level expectations!

> **If so, volunteer for further pilot testing!**

To Scale or Not To Scale?

➢ **If standards cover different content from one grade to the next, vertical scaling may not be advisable!**

- ▪ Inferences based on the assessment at one grade may not generalize well to achievement of the next grade's content.
- ▪ It may still be reasonable to ask whether students are making progress getting closer to (or further above) expectations for each grade.

➢ **If vertical content is well-articulated across grades, it may be reasonable to develop a vertical scale.**

Vertical Scaling 101

➢ What is a Vertical Scale?

  ▪ A numerical scale that links or spans multiple levels (frequently grade levels) of a test

➢ Potential Scale Properties (Scale is usually arbitrary)

  ▪ Ordinal (greater is greater and less is less)

    ✓ Sally's score on the 3rd grade assessment is higher than Henry's score on the 4th grade assessment

  ▪ Interval (can compare differences at different scale points)

    ✓ Henry Scored 31 in the 3rd grade and 42 in the 4th grade

    ✓ Sally scored 43 in the 3rd grade and 49 in the 4th grade

    ✓ Henry's score increased more from 3rd to 4th grade than Sally's

  ▪ Ratio (Ratio's have meaning)

    ✓ Henry knew only about 3/4 as much in the 3rd grade as he did in the 4th grade.

Example of a Vertical Scale
(Displayed Horizontally)
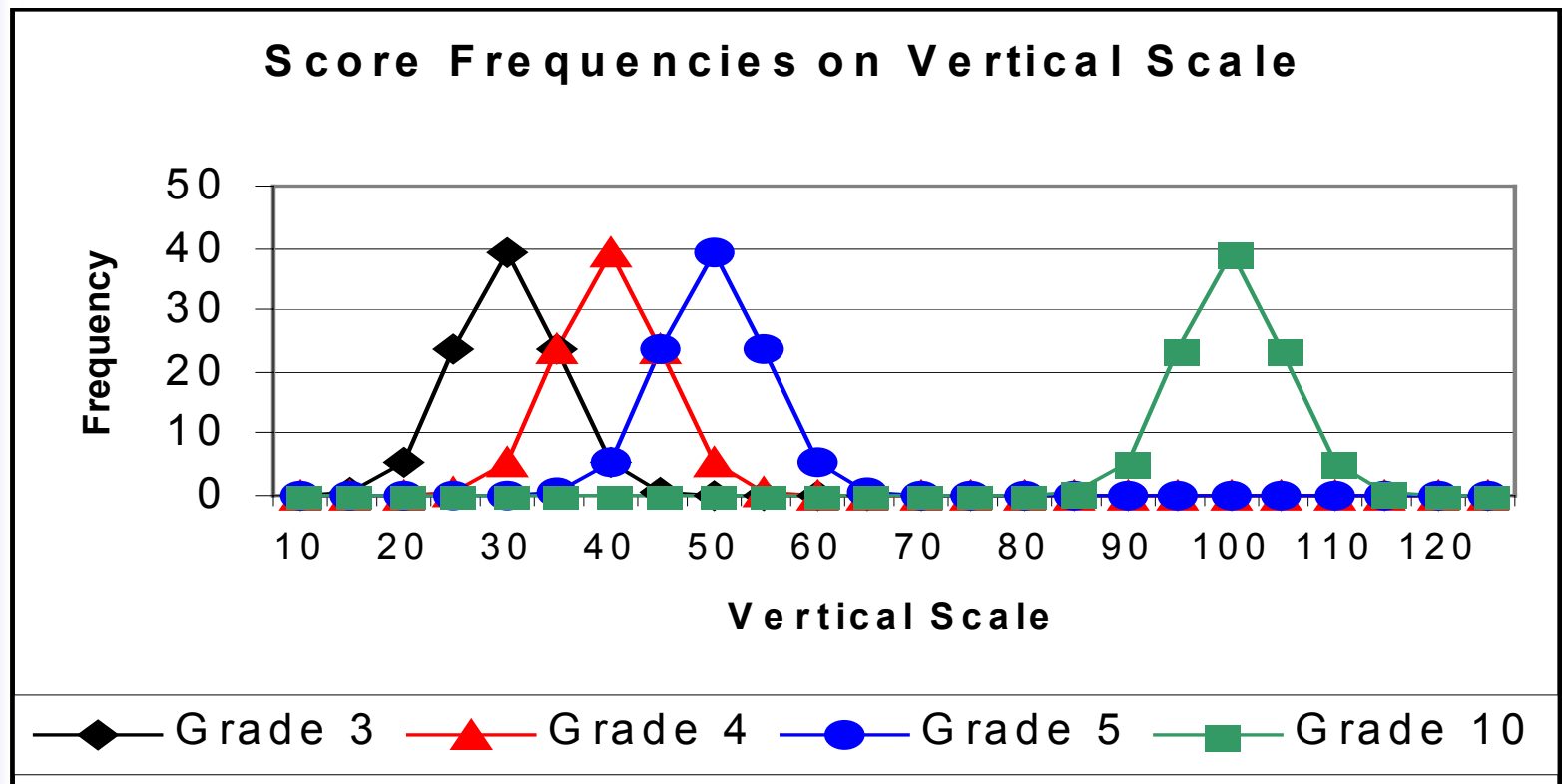
# *What are Vertical Scales Good For?*

Among many possible uses are:

➢ Charting individual student growth across grades

- Identify students whose growth is below expectation and help them
- Growth versus level perspectives

➢ Evaluating specific programs

- Vertical Scales may be a poor substitute for pre- and post-test designs (where both tests measure the same thing)

➢ Assessing "Value-Added" at a particular grade

- Generally defined in terms of average "growth" across students
- May provide fairer comparison of schools serving students at different levels of readiness

How are Vertical Scales Constructed?

➢ **By fiat (Definition): "Grade Equivalent Scales"**

- Scales for each grade are normalized (Normal Curve Equivalents)
- Means are scaled to be 1 unit (grade) apart
- So 3.3 could mean 3/10th standard deviation above the grade 3 mean or 7/10th standard deviation below the grade 4 mean
- No checks that the scales measure the same thing
  - ✓ Betsy was below grade-level for reading in the 3rd grade, but above grade level for throwing a football in grade 4.

➢ **By "Equating" the scales used for different grades**

- Some students take both assessments
- Students at each grade take a common set of (anchor) items
- Can provide evidence of the extent to which the assessments being linked measure the same thing.

Sample Design for Vertical Equating

Sample Configuration of Items on Vertical Scaling Test Forms
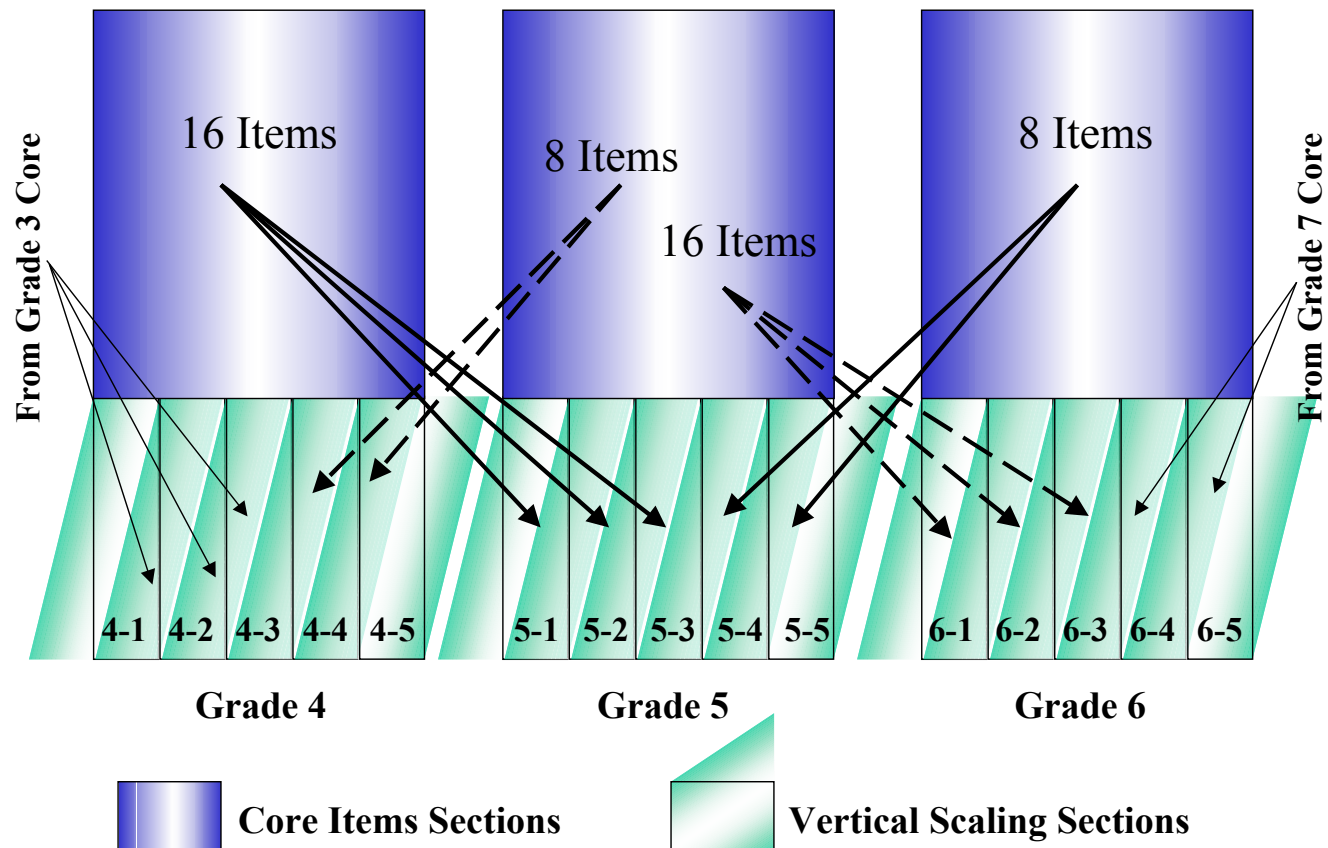
Useful References for Vertical Scaling

Specific to Vertical Scaling

➤ Schulz, E.M. & Nicewander, A. (1997). Grade equivalent and IRT representations of growth. *Journal of Educational Measurement*, 34(4), 315-332.

➤ Yen, W. M. & Burket, G.R. (1997). Comparison of item response theory and Thurstone methods of vertical scaling. *Journal of Educational Measurement*, 34(4), 293-314.

➤ Camilli, G. (1999). Measurement error, multidimensionality, and scale shrinkage: a reply to Yen and Burket. *Journal of Educational Measurement*, 36(1), 73-78.

➤ Williams, V.S.L., Pommerich, M. & Thissen, D. (1998). A comparison of developmental scales based on Thurstone methods and item response theory. *Journal of Educational Measurement, 35(2)*, 93-107.

More General References on Equating

➤ Kolen, M.J. & Brennan, R.L. (1995). *Test Equating Methods and Practices.* New York: Springer.

➤ Peterson, N.S., Kolen, M.J., & Hoover, H.D. (1989). Scaling, norming and equating. In Linn, R.L. (Ed.) *Educational Measurement, 3rd Edition.* New York: American Council on Education and Macmillan.

➤ Lord, F.M. (1980). *Applications of Item Response Theory to Practical testing Problems.* Hillsdale, NJ: Lawrence Erlbaum Associates.