



**ESU#13 NWEA Achievement Levels Test  
(ALT or Levels-paper and pencil version)  
and  
Measures of Academic Progress  
(MAP-computerized adaptive version)  
-Assessment Description and Rational for Design:  
  
ASSESSMENT DEVELOPMENT**

The Educational Service Unit #13 Consortium Math Achievement Level Test, (ALT) is locally referred to as “NWEA Levels.”

The NWEA Levels Growth criterion-referenced tests were developed through a consortium of schools with Educational Service Unit #13 and the North West Evaluation Association of Portland, Oregon. NWEA Levels Growth tests were chosen because of their unique ability to contribute to the assessment program of schools varying in size from Class I (small, very rural schools often with less than ten students) to Class III schools (as large as 3,000 students). These tests allow:

- the district to assess achievement on standards,
- to give each student a test appropriate to his or her achievement level
- to more accurately measure learning growth from year to year.

There are multiple levels of each test. Each level contains items that are more difficult than those of the preceding level as well as items easier than those at the next higher level. Meaningful growth for each student can then be measured regardless of the starting point.

These tests use a specially designed calibrated scoring system called a RIT scale (Rasch Unit) that ensures each student has the opportunity to succeed and maintain a positive attitude toward testing. Focusing the tests made it possible to develop shorter tests, use less class time and still maintain high reliability.

The difficulty range of the items in each level overlaps the levels immediately above and below it in the series. A student is assigned a test where the mid-range of difficulty approaches his/her achievement level based upon a “locator” test. If testing scores indicate inaccuracies, students are retested immediately to ensure the most accuracy possible for the assessment system. The district uses normative data supplied by NWEA to determine the percentile at grade 4, 8, and 10 for state reporting of proficiencies.

Because NWEA Levels are continuous level tests, it is possible to have up to eight different paper and pencil tests given to various students in a single classroom. This test portfolio does not contain all 8 levels. Instead, to make this report manageable, a statistical study was completed to



determine which test forms were taken most often by the students at this grade level. The tests below were used for post development review panels and actual test samples are included in this portfolio.

Educational Service Unit 13 Consortium Math Achievement Level Test Statistical Rationale for Selection of the Sample Tests for Post-Review Study			
Test Form	Level	% students taking this form	Total percentage of grade level students represented by these two forms
Math 4 <sup>th</sup>	B04	28.2%	51.8%
	B05	23.6%	
Math 8 <sup>th</sup>	C07	18.7 %	44.1%
	C08	25.4%	

Levels Paper and Pencil Test is Equivalent to the Computerized Adaptive Version called MAPS

When the Levels Consortium wrote their Goals 2000 Grant to develop Levels tests, they intended to eventually develop a computerized adaptive version. Adaptive tests tend to be more challenging to high performing students and less frustrating to low performing students. In every research study done to date, the evidence indicates that students are, on the average, more motivated by adaptive tests than by conventional paper and pencil tests or by conventional tests administered by computer. No evidence of bias by ethnic group or gender has been noted to date. In the fall of 2000, NWEA finally had this option available for the consortium schools. The title of the computerized version is “MAP: Measures of Academic Progress.”

The MAP is a survey test with goal scores identical to the paper and pencil goal scores. Both are designed to return a very accurate overall RIT score in a content area plus goal performance for the major goals in the content area. The NWEA computerized item data base has been aligned to the goal structures determined by the original paper and pencil test. Students taking the computerized test will get some of the same items as students on the paper and pencil. Additional item from the test data bank, which calibrate to the same goals, are occasionally selected by the computer in order to provide the needed the needed range and variety for students who take the test more than once. The computerized version also has the fall items from which to pull. The MAP test is statistically equivalent to the Levels test because :

- they are designed on the same goal structure
- computerized test items were calibrated with the paper and pencil test items
- another consortium, who used the same developmental process, calibration, and test item data bank, conducted a formal study of equivalence.
  - A Comparison of MAP and ALT Scores, Executive Summary; dated 3/2/2001, is included in the Levels Test Appendix following this divider.
  - The study compared students on both Achievement Level Tests and Computerized Measures of Academic Progress. Correlations between Spring ALT scores and Fall ALT scores range from .88-.90. Correlations between Spring ALT scores and Fall MAP scores range from .83-.85. The test-retest correlations observed on both cases are extremely high, and those observed in changing test modes from ALT to MAP are only fractionally lower than those observed when the test mode remains constant.



While a student used a pencil and special answer sheet to take an achievement level test, the student takes a computerized adaptive test on a computer. A test administrator usually sets up and monitors the test. When each question is displayed on the computer screen, the student selects an answer using the mouse or keyboard.

Like the paper and pencil test, the whole class does not receive the same computerized test. The paper and pencil tests are preceded by short Locator Tests designed to identify and then place each student at the appropriate ability and testing level. This ensures that the majority of the test questions are at the right difficulty level to be able to get accurate growth scores. The MAP can adjust to the student's performance, so the Locator Test is unnecessary. Within an optimum test, the student answers approximately half the item wrong and half the items right.

2000-2001 was the first year for some consortium members to use the MAP instead of the equivalent Level paper and pencil test. Results on either test are interchangeable for district decision-making.

The ALT (Levels) paper and pencil test has been used to complete the Quality Criteria goals that follow.



***QUALITY CRITERION ONE:***  
***The assessments reflect the state or local standards.***

**GRADE LEVEL 8**

***1. A representative panel of qualified teachers or other educators from the district has judged the assessments to be: (a) matched to the standards, and (b) adequate to cover the standards.***

Consortium Documentation

**Panel 1: Original Assessment Development Committees**

The Assessment Development Committee and the districts they represent are listed below. Since the initial test design, additional districts have joined the consortium and now receive assessment services directly from NWEA, using this same personalized assessment battery. In the summer of 1999, state standards were not mandatory. The ESU#13 Consortium was formed to meet local needs to assess the standards they deemed most important and which were also possible to assess in a selected response format. The NWEA company was selected primarily because of their quality reputation in helping schools develop criterion referenced tests which allowed individualized student testing and essential growth scores.

In late 1999, the state of Nebraska indicated requirements to test all standards. The Consortium desired to continue to use the “Levels” tests they had developed for regional needs and to supplement it with additional assessments. They then developed more comprehensive assessment systems involving another norm referenced test, the Levels, and a regionally developed criterion referenced test called ESU#13 STARS. In addition, informal classroom assessments are used, but not for state reporting. Together, the system provides a comprehensive picture of achievement progress and allows each assessment to contribute its best attributes.

<b>ESU#13 Consortium NWEA LEVELS Assessment Design Team, 1998</b>								
Participants	Years of Experience in current district= <b>15.8 years average for group</b>	Districts	Size: Large (Total student FTE > 1000 students)	Size: Small (Total district FTE < 1000 students)	Role: Classroom Teacher/Substitute Teacher	Role: Curriculum Specialist	Specialty	Experience in working with special needs: poverty, migrant, title, special education
<b>Math Totals: 18</b>	<b>284</b>	<b>12</b>	<b>4</b>	<b>11</b>	<b>13</b>	<b>5</b>	<b>K-12 expertise</b>	<b>18</b>

Educational Service Unit 13 MATH Achievement Level Tests: Spring MATH  
Criterion One: Alignment



Michelle Peters	3	Mitchell EL		x	X		Title I remedial math	X
Deanna Keszler	16	Scottsbluff EL	x		X		3-4 grade teacher	X
Lisa Blanco	14	Haig EL		x	X		3-4 grade teacher	X
Carolyn Haeffelin	10	Bridgeport EL		X	X		4 <sup>th</sup> grade teacher	X
Karen French	26	Mitchell EL		X	X		4 <sup>th</sup> grade teacher	X
Ly nne Goos	32	Gering EL	X		X		4 <sup>th</sup> grade teacher	X
Todd Hastings	8	Chadron MS		X	X		5 <sup>th</sup> -6 <sup>th</sup> grade teacher	X
Nan Kissler	12	Minatare EL		X	X		6 <sup>th</sup> grade teacher	X
Jill Balcom	2	Sioux Co EL		X	X		k-8 teacher	X
Connie McDaniels	34	Scottsbluff MS	X		X		7 <sup>th</sup> grade math teacher	X
Bob Cooper	24	Banner Co HS		X	X		7-12 Math teacher	X
Matthew Chrisman	4	Rushville HS		X	X		9-12 Math teacher	X
Linda Drinkwalter	7	Chadron HS		X	X		9-12 Math teacher	X
Joe Baker*	34	Scottsbluff	X			X	Director of Curriculum, Secondary Math Endorsement	X
Myron Lembke*	18	Bridgeport		X		X	Elementary Principal	X
Ron Sylvester*	20+	Consultant/	NA	NA		X	Grant Coordinator	X
Penny Businga* Became regional contact	20	ESU 13	NA	NA		X	Staff and Curriculum Director, ESU wide Endorsed ESL and High Ability	X
Dr. Bill Mengel facilitator	NA	Consultant	NA	NA		X	Curriculum Specialist and Consultant for NWEA test development	X

#### Levels Assessment Development Committee

- 16 members developed the test

#### Process used by Levels Test Development Committee

##### Steps Followed:

1. Regional educators with a wide variety of backgrounds and education were trained in understanding standards by ESU #13 staff developers. This training had each participating district do 5 steps:
  - discuss the deep meaning of each standard for content and for behavior required
  - align the state standards to the district curriculum
  - share with other districts in the consortium to compare meanings and alignment
  - develop a final common “vocabulary” of meanings for each standard and common curriculum for assessment.
  - the consortium mailed copies of curriculum and standards documents to the non-profit assessment company, Northwest Evaluation Association, NWEA
2. NWEA hired a consultant to do the initial review of the state standards and our curriculum in relationship to the NWEA goal structures and sub-goals.
3. Copies of “goal structures” were mailed to consortium, were approval and needed changes were noted.
4. NWEA then made necessary changes to goal outlines.
5. Consortium selected teachers from districts representative grades: K – 12. The tests would be specifically designed for grades 3-8 and high school algebra and geometry, but we wanted to have the total developmental perspective as well as increase the sequential knowledge of participating districts.



6. The NWEA facilitator provided overview training on the Rausch scaling system and how the tests were designed, purpose and use of the data for measuring student growth. He also loaded software onto ESU computers with item banks for each subject. The item banks are the result of ongoing NWEA research, which calibrates the difficulty, assigns RIT values, and provides statistical data. When choosing items, the developers would search the bank for keywords in the standard. Each item was assigned a difficulty level. At the end, graphs which displayed the range of items and their difficulty. Length and test difficulty were adjusted accordingly.
7. Test development training for teachers (June 1999)
  - Day 1 – background information presented  
demonstration lessons of software given  
specifics of test design presented  
information on how to create a fair and balanced test
  - Day 2 – teachers began writing test, creating a fall series and a spring series of tests.
  - Days 3 –4, - teachers finished the tests and reviewed for bias and content. At least two teachers reviewed each level of the test. After test development NWEA Test Production team proofed copies of all tests.
8. Proof tests mailed to consortium who then developed a review team which made any necessary changes and returned them NWEA
9. NWEA Test Production team created print masters of all tests and returned them to the consortium for printing, distribution and administration
10. Tests were developed, proofed, and revised. The tests were initially administered to pilot groups of students in order to develop better testing administration procedures before large scale testing was attempted.(October, 1999)
11. Large scale testing was conducted and additional districts joined the consortium. (April, 2000)
12. The consortium conducted four different intensive trainings on interpretation of scoring results so that teachers would be able to implement appropriate interventions if necessary.
13. The State of Nebraska announced that assessments of standards would be required. (2000)
14. A review team reevaluated the Levels test for its alignment with the newly revised state standards and as a post-development quality control committee. (2001)

Results of Panel 1: Original Assessment Development Committee

- After many revisions, pilot tests were published and administered
- A list of standards were assessed by each subtest were distributed so that teachers could use that specific information for classroom interventions.



- Determined to use composite scores for reporting. Sub-scores are intended to be used to help classroom instruction.
- The original item-by-item alignment to the NWEA goal structures document follows the sample test in this section.. The goal structures were translated into the standards shown on this chart.

NWEA LEVELS Test Development Committee Alignment of Test with Nebraska Standards		
Test Section	Format	Nebraska Math standard met
Math (grades 3-8)	Multiple choice	4.1.3 8.1.3
		4.2.3 8.1.4
		4.3.1 8.2.1
		4.4.3 8.3.1
		4.5.1 8.4.3
		4.6.1 8.5.2
		8.6.2

- The test was used as designed since 1999. However in 2002, an outside review panel reevaluated the test item alignment. The consortium decided to use these new recommendations as the official alignment for state reporting. Longitudinal information is not affected, but some standards did change. See Panel 2 information.

Design Committee’s Involvement in Developing the Assessment

- ESU #13 7 county region has less than 400 teachers of mathematics K-12
- 16 educators (teachers, curriculum specialists) participated in sessions; this is approximately 4 % of the classroom content area teachers possible.
- Hundreds of additional teachers participated in workshops to teach them how to prepare for the tests and how to use the test results.

Time of Year

- Panel 1 was primarily involved in reviewing the assessment for alignment to standards during test development. The test development process involved several quality control checks, including a steering committee, a revision committee, administrative review, and final panel review before pilot publication.
- The tests have been in use since 1999. Most districts have used both fall and spring forms of the test to obtain the desired student growth information. Only the spring version is addressed in this portfolio, since those are the scores used as part of the state reporting process.

Forms

- Forms used by this committee were supplied by NWEA and are no longer available to include as evidence. Committees worked directly on the computer selecting items, monitoring the difficulty level of each and balancing the test for content and difficulty.



**2. A panel of individuals from outside the district has judged the assessments to be: (a) matched to the standards, and (b) adequate to cover the standards.**

Consortium:

**Panel 2: Post-Development Outside Review Panel**

The Post-Development Review Panel was assembled to provide an objective outside review of the Levels test. The panel was designed to represent the region in several ways that are indicated across the top of the panel matrix. Because it is difficult in a very rural area to truly represent all constituents, and because not all invited participants were able to attend, panel members were reminded to use the best of their expertise to make decisions for all children in the region.

Committee Demographics

Participants and their qualifications are shown on the chart below. The totals are summarized just under the headings for each area.

**Post-Development Panel to Review Most Frequently Used NWEA Levels at Grade 4 & 8**

Participant	Years of Experience in Education 24 years average for group	Current Position	Focus	Gender: Male	Gender: Female	Ethnicity: Asian American	Ethnicity: Native American	Ethnicity: Hispanic	Ethnicity: European	Education: Some College/Military training in testing	Education: College graduate	Education: Graduate Degree/s	Role: Classroom Teacher/Substitute Teacher	Role: Curriculum Specialist	Teacher of Special Ed	Experience in working with special needs: poverty, migrant, title, special education
<b>Totals:</b>				<b>7</b>	<b>2</b>	<b>1</b>			<b>8</b>		<b>9</b>	<b>7</b>	<b>5</b>	<b>6</b>	<b>1</b>	<b>9</b>
Brett Gosnell	4	Teacher	Math major jr high	x					x		x	+	x			x
Jim Hawk	30	Retired teacher	Math major high school	x					x		x	MS	x			x
Mary Jo Hoffman	20	Principal	Elementary		x				x		x	MS	Ad	x		x
Kurt Holscher	9	Teacher	Math major high school	x					x		x	+18				x
Joan Kollars	32	Teacher	Math minor elementary		x				x		x	EdS	Ad	x	x	x
Marc McClanahan	38	Retired curr.dir	Elementary	x					x		x	MS		x		x
Gary Schmucker	32	Retired supt.	High School Music	x					x		x	EdS		x		x
Lewis Shoff	40	Retired supt	High School Sci	x					x		x	EdS		x		x
Robert Stack	11	College Professor	Math major higher ed	x		x					x	PhD	x	x		x

- Role:



The Post-Development Review Panel met on March 11 and 12, 2002. They completed an evaluation of the ESU #13 STARS, the NWEA Levels, the ITBS, Terra Nova, Stanford, and CAT 6 tests for several of the Nebraska Quality Indicators.

Process Used by Post-Assessment Review Panel in Reviewing NWEA Levels Math Assessment

- Since it was used by a large number of rural districts, the outside review panel approach was used to obtain information, which could be used by all districts.
- Participants were reminded of and discussed common characteristics appropriate to grade level students in learning development.

Steps followed

1. ESU staff developers outlined a list of potential panel members. Emphasis was placed on ethnicity, experience and geographic location. Invitations were sent to over 25 people. Native American, African American and Hispanic math experts who had not previously served on test development were unavailable in this region. All participants had experience working with special needs students.
2. Panel members attended a short workshop on standards including both content and behavior, facilitated by Penny Businga, Staff Developer at ESU#13.
3. Participants were given forms that required them to make individual determination if the item was the appropriate standard.
4. Then they indicated on the provided form “yes”, it is appropriate, or “no”, it is not appropriate. The form also included other Quality Indicators on which participants voted after further training. They were also given notepaper to record additional comments and identified errors. The facilitator took the group through training and scoring for all the Quality Indicators for each subsection of the test.
5. Each test item was evaluated by at least 3 panel members. A fourth evaluator compiled the information from the previous ratings and made final determination if there was any disagreement.

Final Post-Development Panelists’ Involvement in Developing the Assessment

- None of the Panel members were involved in developing the assessment.

Time of Year

- Post-development Panel was only involved in reviewing the assessment for alignment to standards after test development and piloting. The group met in March 2002.

Forms

The document used by the Final Post-Development Review Panel sample follows:

**ESU 13 Validation Form for Assessments**



Panel 1 Curriculum/Assessment Design Committee		Panel 2 Pre-Pilot Post-Development Review Panel		Panel 3 Final Post-Development Review Panel					
<b>Quality Indicators</b>									
NWEA Levels Math Grade _____ Part ____	NE LEARNS Standard Write in state standard(s) measured by item	Alignment with Standards		Free from Bias				Appropriate Level	
		Meets desired Content Yes or No	Meets desired behavior Yes or No	Item is bias free in each of the following categories  Yes or No				Cognitively Developmental  Yes or No	Reading Level  Yes or No
Item #				Ethnic	Gender	SES	Format		
1									
2									

Results and Actions of Panel: Final Post-Development Review (Related to Alignment))

- Three panel members agreed with each standard assigned to each item. On this test they were allowed to list multiple standards for an item, if strong agreement existed in all such standards.
- Panel reviewed the work of the original design team and compared their standards recommendations to that of the original team. Those in bold matched both design and post review committee. Since the number of items is 5 or greater, the team decided to follow the recommendations of the Post-Design team for state reporting of standards.

Original team

8.1.3

8.2.1

8.3.1

(The original design team had written a general test for the “strand” and then selected a single standard which best covered all aspects of that strand, usually the standard involved with problem solving.)

8.4.2

8.4.3

usually

8.5.2

8.6.2

Post Design Team recommendations

Suggested that all math standards could be listed if the Levels test was also backed up with the ESU#13 Stars test. See the ledger sheet.

- The NWEA Levels Test consortium is scheduled to redesign the framework for the computerized version in summer of 2002.



### Item Analysis of ESU 13 Consortium NWEA Levels Math Tests to Nebraska State Standards/Regional Curriculum Standards

Regional Curriculum Standard	Numeration				Computation/Estimation					Measurement		Geometry						Data Analysis				Algebraic Expression		
	8.1.1	8.1.2	8.1.3	8.1.4	8.2.1	8.2.2	8.2.3	8.2.4	8.2.5	8.3.1	8.3.2	8.4.1	8.4.2	8.4.3	8.4.4	8.4.5	8.4.6	8.5.1	8.5.2	8.5.3	8.5.4	8.6.1	8.6.2	8.6.3
NWEA Levels Growth CO7 Spring  Taken by approx 19% of 8 <sup>th</sup> graders  <i>(Test item numbers are recorded in the section below the standard they assess as approved by three outside evaluators)</i>	46 47	40 45 46		36	48 21	1 4 5 12 16 21 30 42	1 4 5 30 41 49	11 28 43	5 31	12 21 38 42	20 31 35	6 8 13 26 34 44 50		10 14 39	48		26 50	19 33	2 3 9 15 16 19 22 24 29	17 18 32	18	3 34	7 16 23 25 27 28 37 43 49	9 15 22
<b>Total Items per Standard</b>	<b>3</b>	<b>3</b>	<b>0</b>	<b>1</b>	<b>2</b>	<b>8</b>	<b>7</b>	<b>3</b>	<b>2</b>	<b>4</b>	<b>3</b>	<b>7</b>	<b>0</b>	<b>3</b>	<b>1</b>	<b>0</b>	<b>2</b>	<b>2</b>	<b>9</b>	<b>3</b>	<b>1</b>	<b>2</b>	<b>9</b>	<b>3</b>
<b>Total items per Strand</b>	<b>Numeration items: 7</b>				<b>Computation/Estimation items: 22</b>					<b>Measurement items: 7</b>		<b>Geometry items: 13</b>						<b>Data Analysis items: 15</b>				<b>Algebraic Expression items: 14</b>		
NWEA Levels Growth CO8 Spring  Taken by approx 25% of 8 <sup>th</sup> graders  <i>(Test item numbers are recorded in the section below the standard they assess as approved by three outside evaluators)</i>	43 44	44 50		16	15 29 34 40 41 45 49	13 19 24 49	12 13 15 27 30 31 33 34 37 46 48 50	9	11 42	22 19	2 11 18	7 17 25 47	36	28 35 38 39 40	26 44	36	7 47	3 14 37	3 4 8 21	1 12 33	1	17	5 6 9 10 20 24 28 32 35 36	4 21
<b>Total Items per Standard</b>	<b>2</b>	<b>2</b>	<b>0</b>	<b>1</b>	<b>7</b>	<b>4</b>	<b>12</b>	<b>1</b>	<b>2</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>1</b>	<b>5</b>	<b>2</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>3</b>	<b>1</b>	<b>1</b>	<b>10</b>	<b>2</b>
<b>Total items per Strand</b>	<b>Numeration items: 5</b>				<b>Computation/Estimation items: 26</b>					<b>Measurement items: 5</b>		<b>Geometry items: 17</b>						<b>Data Analysis items: 11</b>				<b>Algebraic Expression items: 13</b>		

**Data Analysis by Final Post-Development Panel:**

- nearly half of 8<sup>th</sup> graders take one of these two level tests, making these levels appropriate for item analysis on 8<sup>th</sup> grade standards
- the most accurate score provided from this assessment is the national percentile based upon total test composite score
- NWEA test users in the ESU#13 Consortium also give the STARS tests which creates additional tests items for consideration for each standard.
- a study of the math standards noted the interrelated nature of the skills indicating that success in a strand suggests ability in the associated standards
- the number of items per strand is adequate to assess students' basic math knowledge in that strand
- because each strand is adequately covered, the panel believes that the use of the total composite %ile score can be used within the district tracking system (to be considered with the STAR test) for all Nebraska standards
- this assessment can be used for reporting student progress on all Nebraska math standards when used in conjunction with the ESU 13 STARS Math Assessment
- if used in conjunction with ESU 13 STARS, translate %ile to the following state proficiency levels :0-24=Beginning, 25-49=Progressing, 50-74=Proficient, 75-100=Advanced for the following math standards:  
8.1.1, 8.1.2, 8.1.3, 8.1.4, 8.2.1, 8.2.2, 8.2.3, 8.2.4, 8.2.5, 8.3.1, 8.3.2, 8.4.1, 8.4.2, 8.4.3, 8.4.4, 8.4.5, 8.4.6, 8.5.1, 8.5.2, 8.5.3, 8.5.4, 8.6.1, 8.6.2, 8.6.3,
- if the NWEA test is the only one used , translate %ile to the following state proficiency levels :0-24=Beginning, 25-49=Progressing, 50-74=Proficient, 75-100=Advanced for the following math standards only:  
8.2.3, 8.6.2 (The Buros study published in 2001, based upon incomplete data, identifies standards 8.2.2 and 8.4.1)
- the panel suggests that this assessment tool not be used alone to determine student performance on individual standards
- this assessment provides statistically valid information to measure individual student progress in mathematics as described in value-added research



**Results and Action Taken to Rectify**

Action 1: In July, 2002, a NWEA Assessment Revision Committee was convened to review this data and suggest changes to make the electronic version (MAPS).

- It is even more closely aligned than earlier versions.

**A. Description of Assessment Revision Committee**

- The Assessment Revision Committee is described in the table below
- Summary of Revision Committee characteristics
  - 60% of Assessment Revision had helped develop original NWEA Assessments
  - All Curriculum Coordinators in the region were involved

Participants	Years of Experience in Education <b>18.6</b> years average for group	School	Worked on original NWEA Assessment Development Committee.	Size: Large (Total student FTE > 1000 students)	Size: Small (Total district FTE < 1000 students)	Location: North	Location: South	Gender: Male	Gender: Female	Education: College graduate	Education: Graduate Degree/s	Role: Classroom Teacher/Substitute Teacher	Role: Curriculum Specialist	Experience in working with special needs: poverty, migrant, title, special education
<b>Totals:</b>				<b>3</b>	<b>4</b>	<b>3</b>	<b>4</b>	<b>2</b>	<b>8</b>	<b>10</b>	<b>8</b>	<b>2</b>	<b>8</b>	<b>10</b>
Joe Baker	31	Scottsbluff	X	X			X	X		X	EdS		X	X
Penny Businga	26	ESU #13	X	NA					X	X	MS		X	X
Twila Griffiths	23	Banner Co	X		X		X		X	X		X		X
Michelle Lembke	6	ESU #13	X	NA					X	X			X	X
Barb McGee	20	ESU #13	X	NA					X	X	EdS		X	X
Teresa Martin	12	Scottsbluff	X	X			X		X	X	MS	X		X
Joy Mockelman	22	Gering Ad		X			X		X	X	PhD		X	X
MariBeth Moore	11	Chadron			X	X			X	X	MS		X	X
Richard Moore	21	Chadron			X	X		X		X	MS		X	X
Sandy Porter	20	Hemingford			X	X			X	X	MS		X	X

**B. Description of Process Used by Assessment Revision Committee**

- ESU #13 Consortium contacted NWEA to discuss item concerns and the best way to make changes to goal structures. NWEA outlined the process and asked the committee to develop the goal structure titles and to list the types of items that fell under these titles. These types could then be matched to the paper/pencil version and the data bank. NWEA agreed to review their comprehensive data bank, to find items which most closely matched, or were on, the original paper/pencil version. This was done to ensure that items on the electronic version matched the revised goal structures for alignment and adequacy of coverage.
- The revision committee reviewed the regional core common math curriculum documents to select their topics under each standard goal area.



- They refined and revised the original goal structures with additional topics and added a few new goal structures to cover additional standards.
- They knew that on the electronic version, 40 items are given for math. Those items are evenly distributed among the goal structures, so having 6 goals in math would result in nearly 7 items per goal structure given to each student.
- The committee contacted NWEA to see if they had enough quality test items which met the alignment requirement and the ability levels.
- Schools were contacted in September with the new goal structures for NWEA MAPS electronic assessments. The realignment resulted in coverage of more of the Nebraska standards, most notably, the high school. (Previous years not enough high school items or students had been involved to make it valid.) This was accomplished by knowingly allowing the electronic version to give students fewer items of each type. Since schools use multiple assessments, the coverage was still seen as adequate to cover Nebraska’s learning standards as shown below:

Math NWEA ESU#13 Consortium Tests			
Original Goal Structures	Alignment to NE Standards	REVISED Math GOAL Structures	Alignment to NE Standards
Use composite (combined) percentile score for state consideration of all listed standards. Use goal structure scores for individualization for each student.			

Goal 1	4.1.3 8.1.3	<b>Goal 1 Number</b> Describe subsets Ordering, comparing Place Value Fraction, deci, percent, equivalencies Positive, negative Money, value, count, computation, change Expanded, exponential, scientific notation, roots, radicals Number theory (prime, composite, factors, multiples, divisibility, powers, properties, absolute value) Ration, proportion	<b>4.1.1 8.1.1 12.1.1</b> <b>4.1.2 8.1.2 12.1.2</b> <b>4.1.3 8.1.3</b> <b>4.1.4 8.1.4</b> <b>4.1.5 8.1.5</b>
Goal 2	4.2.3 8.2.1	<b>Goal 2 Computation, Including Fractions</b> Addition, subtraction of whole numbers Multiplication, division of whole numbers Addition, subtraction of fractions and decimals Multiplication, division of fractions and decimals All operations with ration, proportion, percent, exponents, scientific notation, absolute value, real number properties, roots.	<b>4.2.1 8.2.1 12.2.1</b> <b>4.2.2 8.2.2 12.2.2</b> <b>4.2.3 8.2.3 12.2.3</b> <b>8.2.4</b> <b>8.2.5</b>
Goal 3	4.3.1 8.3.1	<b>Goal 3 Measurement</b> Estimate, measure, calculate metric length, area, perimeter, mass/weight, volume, capacity Estimate, measure, calculate standard unit length, area, perimeter, mass/weight, volume, capacity Estimate, measure using non-standard units Read, measure time and temperature Decimal concepts with money Conversions within measurement systems Problem solving, reasoning applications Measuring tools, accuracy and precision including angles	<b>4.3.1 8.3.1 12.3.1</b> <b>4.3.2 8.3.2 12.3.2</b> <b>4.3.3</b> <b>4.3.4</b>
Goal 4	4.4.4 8.4.3	<b>Goal 4 Geometry</b> Identify, compare, classify, describe dimensional figures and shapes	<b>4.4.1 8.4.1 12.4.1</b> <b>4.4.2 8.4.2</b>



		Identify, describe points, lines, line segments, rays , angles Congruence, symmetry, similarity, transformations, Pythagorean theorem, coordinate geometry Apply formulas for perimeter, circumference, area, volume for geometric figures	<b>4.4.3</b>	<b>8.4.3</b> <b>8.4.4</b> <b>8.4.5</b> <b>8.4.6</b>	<b>12.4.3</b> <b>12.4.4</b> <b>12.4.5</b> <b>12.4.6</b> <b>12.4.7</b>
Goal 5	4.5.1 8.5.2	<b>Goal 5 Data Analysis, Probability and Statistical Concepts</b> Collect, organize, interpret data in graphs, charts, tables Probability, prediction, outcomes Central tendency, measures of variability , range, frequency, sampling techniques Problem solving, reasoning, applications Generate, interpret regression equations Interpret, formulate normal distributions	<b>4..5.1</b>	<b>8.5.1</b> <b>8.5.2</b> <b>8.5.3</b> <b>8.5.4</b>	<b>12..5.1</b> <b>12.5.2</b> <b>12.5.3</b> <b>12.5.4</b> <b>12.5.5</b> <b>12.5.6</b>
Goal 6	4.6.1 8.6.2	<b>Goal 6 Algebraic Concepts</b> Solve, simplify expressions, equalities, inequalities, systems of equations Identify, extend patterns Use variables to describe patterns and write number sentences Coordinate systems, relations, functions Order of operations/properties Solve polynomials	<b>4.6.1</b> <b>4.6.2</b>	<b>8.6.1</b> <b>8.6.2</b> <b>8.6.3</b>	<b>12.6.1</b> <b>12.6.2</b> <b>12.6.3</b> <b>12.6.4</b>

Action 2:

The NWEA tests are also used as part of meeting the state requirements for administering a Norm Referenced test. The Nebraska Department of Education hires the Buros Institute to do item analysis for NRT's, but the NWEA electronic versions had never been reviewed. The consortium asked the Nebraska Department of Education to include the NWEA tests in the next formal NRT review. On June 1-4, 2004, this review was done. The results will not be published until July 2004. At that time the consortium will revise the standards covered to match the new review.

***3. The assessment includes enough items/sufficient performance demonstrations for adequate coverage to determine whether each standard that is assessed has been met.***

Consortium Documentation

**Number of items and demonstrations:**

The recommendations of the Post Development review team were used for adequate number of items. A minimum of 5 items had to be on the assessment and had to be aligned to the standard to be considered for reporting purposes. In addition, the total composite test score in percentile was used resulting in a stronger cumulative scoring of the accepted math standards.

The ledger sheet inserted one page earlier to show alignment also indicates the total number of items per standard. Remember, the test is scored, for reporting purposes, as a total composite, using the same proficiency level for all the standards.



***QUALITY CRITERION TWO:  
 Students have an opportunity to learn the content.***

***GRADE LEVEL 8***

***See individual district sections of the portfolio.***

***Some other method was used to determine that students have an opportunity to learn the content prior to their being assessed.***

Consortium Documentation

Teacher Training Related to NWEA Levels Test Classroom Interventions

October 5, 1998	Interpreting Test Scores Jim Coyle, Indiana and NWEA	50 participants
November 16, 1999	Interpreting the Data Wayne Neuburger, NWEA	50 participants
May, 2000	Correlating Levels and MAPS Frameworks John Cronin, NWEA	47 participants
September 13, 2000	NWEA Computerized Assessment Implementation	41 participants
February 19, 2001	It's About Time Carolyn Mock, NWEA Indiana	60 participants
October 8,9,23, 2001	Data Analysis w/ Levels	40 participants
October 31-Nov. 1, 2001	Data Analysis w/ Levels	8 participants
November 6-7, 2001	Data Analysis w/ Levels	32 participants
February 18, 2002	Solving NWEA Tech Issues	18 participants
December 18, 2002	Assessment Interpretation NWEA	39 participants
April 7, 2003	Value Added Assessment using Levels	37 participants
Dec. 9, 2003	NWEA Report Site Training	35 participants
Feb 26-27, 2004	NWEA Climbing the Data Ladder: Differentiation	38 participants
Sessions are also planned for:		
August 11-12, 2004	Leading with Data	
October 6-7, 2004	NWEA Climbing Data Ladder: Differentiation Follow-Up	

The sessions focused on the importance of using data to make informed classroom decisions. Administrative participation created the leadership capacity needed for long-term building usage of the skills.



***QUALITY CRITERION THREE:  
The assessments are free from bias or offensive situations.***

***GRADE LEVEL 8***

***1.a. The items were selected from a nationally recognized test bank controlled for fairness.***

Consortium Documentation

NWEA Bias Review

- The Northwest Evaluation Association conducts its own bias review procedures on its test items. They outlined their procedure as follows:
  - Once items are written and edited, an independent panel of teachers who represent a variety of racial/ethnic backgrounds reviews the items. This panel examines each item to identify whether it contains words or passages that do not pertain to the skill being tested, but which might be misleading or misunderstood by students with a particular racial/ethnic background. Items, which have difficulties are edited by the review panel, if possible, or are sent back to the author to be revised. In addition, this panel may suggest wording and name changes that might help the set of items being reviewed to have more diversity. The focus of this panel is to see that all students have a fair opportunity to answer the item correctly, without being distracted or misled by the context of the item.

***1. The assessment writers have participated in an orientation regarding test bias.***

Consortium Documentation

Names and Qualifications of facilitator

Bill Mengel, Ed.S., consultant for Northwest Evaluation Association.

Test bias orientation used

- Facilitator provided training in bias related to test development.
- They provided criteria for constructing tests without bias.
- They gave directions concerning the rewording of test items without skewing the items or norming data.



- The Criteria for Constructing Tests without Bias is included in the Levels Test sample.

Forms

- No forms were used with this group.

**2. A panel of qualified educators or others representing various socio-economic and ethnic groups has reviewed drafts of the assessment.**

Consortium Documentation

**Panel 2: Post-Development Review Panel**

The Post-Development Review Panel was assembled to provide an objective outside review of the Levels test. The panel was designed to represent the region in several ways that are indicated across the top of the panel matrix. Because it is difficult in a very rural area to truly represent all constituents, and because not all invited participants were able to attend, panel members were reminded to use the best of their expertise to make decisions for all children in the region.

Committee Demographics

Participants and their qualifications are shown on the chart below. The totals are summarized just under the headings for each area.

**Post-Development Panel to Review Most Frequently Used NWEA Levels at Grade 4 & 8**

Participant	Years of Experience in Education 24 years average for group	Current Position	Focus	Gender: Male	Gender: Female	Ethnicity: Asian American	Ethnicity: Native American	Ethnicity: Hispanic	Ethnicity: European	Education: Some College/Military training in testing	Education: College graduate	Education: Graduate Degree/s	Role: Classroom Teacher/Substitute Teacher	Role: Curriculum Specialist	Teacher of Special Ed	Experience in working with special needs: poverty, migrant, title, special education
<b>Totals:</b>				<b>7</b>	<b>2</b>	<b>1</b>			<b>8</b>		<b>9</b>	<b>7</b>	<b>5</b>	<b>6</b>	<b>1</b>	<b>9</b>
Brett Gosnell	4	Teacher	Math major jr high	x					x		x	+	x			x
Jim Hawk	30	Retired teacher	Math major high school	x					x		x	MS	x			x
Mary Jo Hoffman	20	Principal	Elementary		x				x		x	MS	Ad	x		x
Kurt Holscher	9	Teacher	Math major high school	x					x		x	+18				x
Joan Kollars	32	Teacher	Math minor elementary		x				x		x	EdS	Ad	x	x	x
Marc McClanahan	38	Retired curr.dir	Elementary	x					x		x	MS		x		x
Gary Schmucker	32	Retired supt.	High School Music	x					x		x	EdS		x		x
Lewis Shoff	40	Retired supt	High School Sci	x					x		x	EdS		x		x
Robert Stack	11	College Professor	Math major higher ed	x		x					x	PhD	x	x		x



Role:

The Post-Development Review Panel met on March 11 and 12, 2002. They completed an evaluation of the ESU #13 STARS, the NWEA Levels, the ITBS, Terra Nova, Stanford, and CAT 6 tests for several of the Nebraska Quality Indicators.

Test bias orientation

- The Post-Development Review Committee participated in a Bias Training developed by Dr. Leslie Lukin of Lincoln Public Schools and facilitated by ESU #13 staff developers who are trained by Dr. Lukin. The handouts, which were developed by Dr. Lukin, showing the overheads used in the orientation, are the same as those used in developing the STARS tests, and copies can be seen that section of the portfolio.

Process Used by Post-Assessment Review Panel in Reviewing NWEA Levels Math Assessment

- Since it was used by a large number of rural districts, the outside review panel approach was used to obtain information, which could be used by all districts.
- Participants were reminded of and discussed common characteristics appropriate to grade level students in learning development.

Steps followed

1. ESU staff developers outlined a list of potential panel members. Emphasis was placed on ethnicity, experience and geographic location. Invitations were sent to over 25 people. Native American, African American and Hispanic math experts who had not previously served on test development were unavailable in this region. All participants had experience working with special needs students.
2. Panel members attended a short workshop on bias facilitated by Penny Businga, Staff Developer at ESU#13.
3. Participants were given forms that required them to make individual determination if the item was biased in some way. After bias orientation, panel members were reminded to make individual determinations.
4. Then they indicated on the provided form “yes”, it is bias free, or “no”, it is not. The form also included other Quality Indicators on which participants voted after further training. They were also given note paper to record additional comments and identified errors. The facilitator took the group through training and scoring for all the Quality Indicators for each subsection of the test.
5. Each test item was evaluated by at least 3 panel members. A fourth evaluator compiled the information from the previous ratings and made final determination if there was any disagreement.

Final Post-Development Panelists’ Involvement in Developing the Assessment

- None of the Panel members were involved in developing the assessment.



Time of Year

- Post-development Panel was only involved in reviewing the assessment for alignment to standards after test development and piloting. The group met in March 2002.

Forms

The document used by the Final Post-Development Review Panel sample follows:

ESU 13 Validation Form for Assessments									
Panel 1 ___ Curriculum/Assessment Design Committee		Panel 2 ___ Pre-Pilot Post-Development Review Panel		Panel 3 ___ Final Post-Development Review Panel					
NWEA Levels Math Grade  Part ___		NE LEARNS Standard		Quality Indicators					
		Write in state standard(s) measured by item		Alignment with Standards		Free from Bias		Appropriate Level	
		Meets desired Content Yes or No	Meets desired behavior Yes or No	Item is bias free in each of the following categories  Yes or No		Cognitively Developmental  Yes or No	Reading Level  Yes or No		
Item #				Ethnic	Gender	SES	Format		
1									
2									

Results and Actions of Panel: Final Post-Development Review (Related to Alignment)

- Three panel members voted on each criteria, a fourth panel member compiled and was to voted if there was a discrepancy.
- The four bias areas considered were: Ethnicity, Gender, Socio-economic, and Format.
- After panelists rated all items, a large group discussion was held to make commendations, to share concerns, and to propose recommendations
- Results:

CO7: 100% of items were rated with no bias.

CO8 Test: 100% of items were rated with no bias.

- The NWEA Levels Test consortium was scheduled to redesign the framework for the computerized version in summer 2002. Above items were reviewed, as well as the entire process for item selection.



***QUALITY CRITERION FOUR:***  
***The level is appropriate for students.***

**GRADE LEVEL 8**

***Other: A nationally recognized assessment organization provided assistance in designing ability leveled tests to increase appropriateness of test for each student.***

Demonstrated effectiveness

The NWEA item test bank is built using Item Response Theory. This is used to create a measurement scale that is stable across different groups of individuals and has recently been proven to be stable across substantial periods of time. The NWEA item banks are based upon one-parameter logic IRT models also called the Rasch Model. Over time thousands of items have been added to these banks. Each item has been connected to the original measurement scale through the use of IRT procedures and systematic measurement practices. A study conducted in January, 2001 called, The Stability of Measurement Scales: A Longitudinal Study found extremely high correlations between original and new item difficulties. A summary of the study is included in the Levels Test Appendix.

Students assessed in tested ability range

The Achievement Levels Test utilizes six means of ensuring appropriate level.

- the difficulty range of items on the calibrated measurement scale
- test design overlaps difficulty in adjacent levels. The design provides 50% overlap in item difficulty between adjacent levels. This allows students who are misplaced by one level to still obtain scores with a fair amount of accuracy.
- test accuracy is enhanced because all of the test items are within the students ability, not just a few. This results in a fairly narrow range of item difficulty in a particular level. This provides desirable psychological and psychometric consequences.
- the Levels test length stays appropriate while still providing a maximum of appropriate questions
- the use of a Locator test (pretest) to appropriate place students
- retesting of students who score too high or too low for proper measurement



***1. A representative panel of qualified teachers or other educators from the district has judged the assessments from a developmental and cognitive perspective.***

Consortium Documentation

**Panel 1: Assessment Development Committee**

- Complete panel information is outlined in Quality Criterion One.
- Some sample information is repeated below.

Assessment Development Committee:

100% are reading/language arts educators

77% are currently classroom experts

23% are curriculum and developmental specialists

48% have 15 or more years of experience

Process

- Test developers were divided into grade level groups to select the items for the test in each goal and sub goal area. The NWEA test bank has been extensively tested for appropriate level and difficulty. The computer included guidelines and would prompt developers if they had selected too many items of one difficulty.
- Test developers reviewed their current curriculum and other materials to ensure the appropriateness of the selected items.

Results and Actions to rectify

- The revision and editing committee made changes to tasks and items as recommended by the test developers at a quality review session in September, 1998.

Forms

- No forms are available for view because they were included in the software.

***2. A panel of individuals from outside the district has judged the assessments to be of appropriate difficulty.***

Consortium:

**Panel 2: Post-Development Outside Review Panel**

The Post-Development Review Panel was assembled to provide an objective outside review of the Levels test. The panel was designed to represent the region in several ways that are indicated across the top of the panel matrix. Because it is difficult in a very rural area to truly represent all constituents, and because not all invited participants were able to attend, panel members were reminded to use the best of their expertise to make decisions for all children in the region.

Committee Demographics



Participants and their qualifications are shown on the chart below. The totals are summarized just under the headings for each area.

**Post-Development Panel to Review Most Frequently Used NWEA Levels at Grade 4 & 8**

Participant	Years of Experience in Education 24 years average for group	Current Position	Focus	Gender: Male	Gender: Female	Ethnicity: Asian American	Ethnicity: Native American	Ethnicity: Hispanic	Ethnicity: European	Education: Some College/Military training in testing	Education: College graduate	Education: Graduate Degree/s	Role: Classroom Teacher/Substitute Teacher	Role: Curriculum Specialist	Teacher of Special Ed	Experience in working with special needs: poverty, migrant, title, special education
<b>Totals:</b>				<b>7</b>	<b>2</b>	<b>1</b>			<b>8</b>		<b>9</b>	<b>7</b>	<b>5</b>	<b>6</b>	<b>1</b>	<b>9</b>
Brett Gosnell	4	Teacher	Math major jr high	x					x		x	+	x			x
Jim Hawk	30	Retired teacher	Math major high school	x					x		x	MS	x			x
Mary Jo Hoffman	20	Principal	Elementary		x				x		x	MS	Ad	x		x
Kurt Holscher	9	Teacher	Math major high school	x					x		x	+18				x
Joan Kollars	32	Teacher	Math minor elementary		x				x		x	EdS	Ad	x	x	x
Marc McClanahan	38	Retired curr.dir	Elementary	x					x		x	MS		x		x
Gary Schmucker	32	Retired supt.	High School Music	x					x		x	EdS		x		x
Lewis Shoff	40	Retired supt	High School Sci	x					x		x	EdS		x		x
Robert Stack	11	College Professor	Math major higher ed	x		x					x	PhD	x	x		x

- **Role:**  
 The Post-Development Review Panel met on March 11 and 12, 2002. They completed an evaluation of the ESU #13 STARS, the NWEA Levels, the ITBS, Terra Nova, Stanford, and CAT 6 tests for several of the Nebraska Quality Indicators.

Process Used by Post-Assessment Review Panel in Reviewing NWEA Levels Math Assessment

- Since it was used by a large number of rural districts, the outside review panel approach was used to obtain information, which could be used by all districts.
- Participants were reminded of and discussed common characteristics appropriate to grade level students in learning development.

Steps followed

1. ESU staff developers outlined a list of potential panel members. Emphasis was placed on ethnicity, experience and geographic location. Invitations were sent to over 25 people. Native American, African American and Hispanic math experts who had not previously served on test development were unavailable in this



- region. All participants had experience working with special needs students.
2. Panel members attended a short workshop on readability and level of difficulty facilitated by Penny Businga, Staff Developer at ESU#13. The workshop led discussion on cognitive level and readability. Participants were also reminded that the list of accommodations allows a classroom teacher to read the text words to students in special circumstances.
  3. Participants were given forms that required them to make individual determination if the item was the appropriate difficulty.
  4. Then they indicated on the provided form “yes”, it is appropriate, or “no”, it is not appropriate. The form also included other Quality Indicators on which participants voted after further training. They were also given notepaper to record additional comments and identified errors. The facilitator took the group through training and scoring for all the Quality Indicators for each subsection of the test.
  5. Each test item was evaluated by at least 3 panel members. A fourth evaluator compiled the information from the previous ratings and made final determination if there was any disagreement.

Final Post-Development Panelists’ Involvement in Developing the Assessment

- None of the Panel members were involved in developing the assessment.

Time of Year

- Post-development Panel was only involved in reviewing the assessment for alignment to standards after test development and piloting. The group met in March 2002.

Forms

The document used by the Final Post-Development Review Panel sample follows:

ESU 13 Validation Form for Assessments									
Panel 1 ___ Curriculum/Assessment Design Committee			Panel 2 ___ Pre-Pilot Post-Development Review Panel				Panel 3 ___ Final Post-Development Review Panel		
		Quality Indicators							
NWEA Levels Math Grade  Part ___	NE LEARNS Standard Write in state standard(s) measured by item	Alignment with Standards		Free from Bias				Appropriate Level	
		Meets desired Content Yes or No	Meets desired behavior Yes or No	Item is bias free in each of the following categories  Yes or No				Cognitively Developmental  Yes or No	Reading Level  Yes or No
Item #				Ethnic	Gender	SES	Format		
1									
2									



Results and Actions of Panel: Final Post-Development Review

(Related to level of difficulty)

- Three panel members voted on each item. A fourth panel member compiled results. If a discrepancy was found, another panel member was asked to vote.
- . Post Design Team ratings

Levels Math  
Tests C07  
Test CO8  
:all items 100% rated cognitively and developmentally appropriate

- The NWEA Levels Test consortium is scheduled to redesign the framework for the computerized version in summer of 2002. The committee will be aware of the continued need to be diligent in this area.

***3. A readability analysis has been conducted on the assessments.***

Consortium Disclaimer: Because the test items are selected from a statistically calibrated test bank, readability is not an issue. Norming takes the place of readability. In addition, accommodations allow reading a math test to students needing such assistance.



***QUALITY CRITERION FIVE:  
Scoring is consistent.***

***GRADE LEVEL 8***

**Objectively Scored Assessments**

***1. Reliability (consistency) has been estimated by a) a retest, b) an alternative form of the assessment, or c) an internal consistency reliability procedure.***

Consortium Documentation

**Reliability Methods (Marginal Reliability and Mean Classification Reliability)**

Item Response Theory does not use retest, alternative forms, or internal consistency formulas because the test is prone to change as the test sample changes. Individual districts who have developed their tests often do not have the statistical support to complete those methods after development of the test. IRT tests usually use standard error of measurement to be more appropriate because it isn't dependent on the distribution of students that happens to be in a particular class in a particular year.

One approach to estimating reliability which can be modified to work with level testing is the approach of marginal reliability described by Boch and Leiberan (1970). This approach requires that the developer describe the characteristics of the distribution of the trait being measured in the sample being tested. Then, the standard error of measurement is used to calculate reliability through integration across the distribution of the sample. This approach requires the test developer to specify the distribution on the trait in the population of interest and it allows the recalculation of marginal reliability if the population changes in its characteristics.

For the calculation of marginal reliability of the NWEA ALT we assume that the distribution of the trait in the population in each grade and subject is normal. Means and standard deviations used are available upon request. Those sample means were taken from a large norming study in 1996 using students in 37 school districts in 10 different states. The average sample size per grade was slightly over 20,000.

With these means and standard deviations NWEA calculated the marginal reliability in each grade and subject. NWEA assumed that the student takes the most informative level available at the true achievement level. While this assumption is probably not completely accurate, it allows us to estimate reliability without having to estimate the stability of student performance as they pass through the test. Since retest rates in districts using level tests are relatively low (1 to 2 percent) this assumption is



probably not very far from the truth. The consistently high reliabilities are the outcome of the test design and level assignment procedures used in Level testing.

<b>Marginal Reliabilities for NWEA Achievement Levels Test</b>	
Grade	Math Tests
3	.932
4	.935
5	.94
6	.943
7	.947
8	.949
9	.941
10	.930

Another indicator of the reliability of the test scores from the levels series is Mean Classification reliability. Mean Classification reliability identifies the percentage of students that will be correctly classified as being above or below the mean, as a function of the standard error of measurement and the distribution of the population of students. Mean classification reliability is an important indicator of test quality because it indicates how often a high-performing would be mistaken for a low-performing student. Since a host of educational decisions depend on whether or not a student is identified accurately as being above or below the average, the indicator of reliability has educational consequence, as well as statistical meaning.

<b>Mean Classification Reliabilities for NWEA Achievement Levels Test</b>	
Grade	Math Tests
3	.974
4	.974
5	.973
6	.972
7	.970
8	.966
9	.968
10	.958

Norming Information

In 1999, NWEA renormed their data on the test items on over 500,000 students from 100 districts in 18 states. The new norming data did not change the RIT scale, just the reference group. The new reference group is more meaningful to ESU#13 consortium



schools. The consortium uses the Levels tests because it allows the comparison of the performance of a single student, school or school district to a larger, meaningful reference group. The RIT scale is designed to provide direct comparisons to the curriculum and standards. However, it can also be used in a norm referenced manner to create scores like percentile ranks.

The process of producing normative information on the NWEA level tests is somewhat different from typical large-scale norming procedures. In those cases, there is usually a single form of a test that is administered to a selected population, or two forms of the test are administered. From that data several types of scores are derived.

In the NWEA scales, each district has constructed a different series of level tests by selecting items from the NWEA Rasch-calibrated banks. Since each item has been calibrated to a common measurement scale using item response theory, scores from different test forms are able to have essentially the same meaning. The tests used by the NWEA staff for the norming study were built under the direct supervision of NWEA staff and are similar in length and content. The standard error of measurement is closely controlled in the testing blueprint, and students are administered a level to maximize the information in the test score. See the *NWEA Achievement Level Test Technical Description* for a discussion of measurement aspects of the NWEA level tests, which is available from the publisher.

The NWEA level test norms are not based on samples of students drawn to match national demographic patterns, but rather include scores from all participating NWEA member districts. This accurately reflects a population of interest of member districts that actively use level tests for making instructional and programmatic educational decisions. The districts used in the sample have been using the level tests for varying lengths of time, ranging from one testing session to over twenty years. States surrounding Nebraska with similar demographics and rural nature (i.e. Colorado, Wyoming, Iowa) were included.

For the Spring math samples:

The grade 4 sample size was 52,580 students.

The grade 8 sample size was 43,093 students

The grade 9 sample size was 16,725 students

The grade 10 sample size was 5,583 students

(For comparison, the National Assessment of Educational Progress NAEP produces outcome norms, which would also not appropriately be called national norms, even though the sampling design is precise. Sample sizes for NAEP were approximately 7,700-11000 students per grade. The TIMSS study of students in 45 countries used 500,000 students, just slightly smaller than the NWEA norm study. Most wide-range achievement tests use student samples of less than 3,000 students per grade level, and many have fewer.)



## ***2. Other : Use of single composite score to decrease measurement error.***

### Consortium Documentation

Recording student proficiencies in Mathematics on the NWEA Levels Tests will be based on a single composite score rather than as “proficiencies” in separate goal areas. This choice was made in consideration of the following factors:

- The total scores are more dependable than those based on a small subset of questions (goal scores) from the complete tests. The total score for a content area is based on four to six times as many test questions as for an individual goal area. In the case of Level Tests, measurement error for total scores are typically in the 2.5 to 4.0 RIT range. Errors associated with goal scores from the same tests tend to range from two to three times the error levels of the total scores. The fewer the questions, the larger the measurement error will be.
- The test questions from a content area have all been empirically placed on a single, uni-dimensional scale based on their difficulty. While the content of the questions can be assigned to a goal area, the goal area scores are inherently inter-related. Goal area scores serve to organize and to inform thinking about the student’s relative strengths and weaknesses. But, by themselves, the information they provide must be treated cautiously.
- Content experts agree that the skills involved in the act doing math are interrelated. Understanding of numeration is essential for all other math tasks. Computation is involved in all areas. Problem solving can overlap. The decision to use and record a composite score recognizes and overcomes the difficulty of separating skills with enough reliability to use in making district decisions.
- Experienced content area experts can disagree about which goal within a content area is being assessed by a particular test item. Assigning test questions to goals is a fairly subjective process. The more important question, in terms of measuring student proficiency in a content area, is whether or not the test question is a part of the content domain of interest. This is determined empirically by examining how well the test question fits the measurement model used to calibrate the other test questions in the pool of questions. Only questions fitting the model become part of the pool. Those not fitting are rejected.



***QUALITY CRITERION SIX***  
***Mastery levels are appropriate.***

***GRADE LEVEL 8***

Discussions on Levels of Performance

Consortium members are concerned about consistency in scoring. The NWEA Levels tests have been approved by the State of Nebraska as usable to meet the requirements for a norm-referenced test as well as for use as a criterion referenced test. Since the Levels test is so considered, the consortium decided to go with the State Department mastery levels established for other normed tests.

***Nebraska State Department of Education published mastery levels related to percentiles.***

Basis for decision

On May 17, 2001 the Nebraska Department of Education held a satellite broadcast for school districts on the reporting procedures and portfolio development and review processes. Part of this broadcast dealt directly with how scores on standardized tests were to be reported and what the cut scores for each level are. The following are the percentiles that scores will be reported. This was determined by the Nebraska Department of Education.

<u>Grade 8</u>	<u>Total Math</u>
75-99%	Number of students at these percentiles
50-74%	Number of students at these percentiles
25-49%	Number of students at these percentiles
1-24%	Number of students at these percentiles

In addition, we are required to report the number of students at this grade level that were not included in the test.

NWEA Level tests are reported in percentiles. Because the scores are reported in percentiles, the consortium has chosen to report the above mastery levels.

Comparison to “Mastery Levels” set for the Levels tests by other state using items from the same data bank:

Since Nebraska elected to set the “pass” or “proficient” level at the 50%ile and above, the ESU#13 Consortium asked NWEA to provide “cut” scores used in other areas on similar items as ours, as a comparison for local discussion.

In 2000, the state of Indiana set their cut score for NWEA test at the 46%ile for Grade 8. The state of Oregon set their cut score for NWEA Levels tests at the 51%ile for Grade 8. In Grade 4, the state of Washington, in 1999, set their cut score at the 79%ile. Oregon used a cut rate of 48% for 3rd grade. Oregon used the 50% at grade 10. Grade 3,8, and 10 were done in the year 2000. This is some indication that the Nebraska scores are not unreasonable.



Final Mastery Level

Consortium districts are reporting final mastery levels based upon several assessment tools. The final determination will be made by teacher judgment based upon all available testing information approved for the district assessment plan. The consortium has determined that teachers will select the highest available score available for state reporting .