

New Mexico Alignment Study

May 20, 2007

Dillard Research Associates

Gerald Tindal

Lindy Crawford

jerry@dillardresearchassociates.com

Introduction

All students in New Mexico are required to demonstrate proficiency on content standards, benchmarks, and grade level performance standards (GLPS's). To accomplish this for students with significant cognitive disabilities, the state has developed Expanded Grade Band Expectations (EGBE's) and linked them to the grade-level performance standards. In addition, the state has developed the New Mexico Alternate Performance Assessment to assess students' progress toward meeting content standards and GLPS's via their link to EGBE's.

The purpose of this report is to evaluate and summarize the: (a) linkage of EGBE's to grade-level performance standards (GLPS's) as well as the cognitive challenge or complexity inherent in the EGBE's, (b) items on the alternate assessment and their degree of alignment to the EGBE's as well their cognitive complexity (depth of knowledge), (c) New Mexico Alternate Performance Assessment holistically using Webb's four dimensions of alignment, and (d) the alignment of the Achievement Level Descriptors (ALD's) to the EGBE's.

This report is the second draft of the alignment study and summarizes the following:

- Section I – Executive summary and highlights of findings
- Section II – Methodology including a summary of the forms and the process used
- Section III – Analysis of the EGBE's including linkage to Grade Level Performance Standards (GLPS's) and Depth of Knowledge (DOK) after reaching consensus.
- Section IV – Item Analysis of the New Mexico Alternate Performance Assessment: (a) alignment to the EGBEs and (b) depth of knowledge (DOK).
- Section V – Summary Statistics of Alignment using Webb's dimensions: (a) categorical concurrence, (b) range of knowledge, (c) depth of knowledge, and (d) balance of representation.
- Section VI – Alignment of Achievement Level Descriptors to EGBE's
- Section VII – Conclusions and Recommendations

Section I – Executive Summary

The New Mexico Performance Assessment was evaluated with respect to its cognitive complexity and the degree to which it aligned with the Extended Grade Band Expectations (EGBE's). In this process, the peer review regulatory guidance was used as an authoritative reference for organizing our study.

For alternate assessments in grades 3 through 8 based on alternate achievement standards, the assessment materials should show a clear link to the content standards for the grade in which the student is enrolled although the grade-level content may be reduced in complexity or modified to reflect pre-requisite skills. For each grade, the State may define one or more alternate achievement standards for proficiency.¹

We first analyzed the degree to which EGBE's were linked to the Grade Level Performance Standards (GLPS's), which are required for general education students. Although the individual judges were quite varied, we used a process that required consensus. We found that the EGBE's were substantially linked to the GLPS's at an appropriate level (2-3 on a 1-4 scale, reflecting appropriate reduction of breadth). We then analyzed the items in the Alternate Performance Assessment in terms of their depth of knowledge and degree of alignment with the EGBE's. Again, we found appropriate variation of depth and substantial alignment overall though some items for specific grade bands and subject areas could use refinement and adjustment to improve either dimension of depth or alignment. We documented that teachers rated the Achievement Level Descriptors (ALD's) as substantially aligned with the EGBE's. Finally, we end the report with a specific set of recommendations.

In these major findings, we also found considerable specific variation in grade bands and subject areas and recommend the New Mexico Department of Education pay particular attention to the numerous tables included in this report and the appendices in order to improve the content-related evidence of its large-scale assessment for students with significant disabilities. We have organized the report in each of these sections and reference specific appendices as appropriate.

¹ U. S. Department of Education Office of Elementary and Secondary Education (April 28, 2004). *Standards and Assessments Peer Review Guidance: Information and Examples for Meeting Requirements of the No Child Left Behind Act of 2001*. Author.

Section II – Methodology

In this section, we describe the materials, the teachers (including demographics, that noted for each subject area, their degree, years of experience, grade level, special education experience, current grade placement, and subject area), and finally, the process for training and reaching consensus.

Materials

All participants were given the following materials (some of which were secure and are not included in this report).

- General Education Assessment Framework
 - Organized by strands, standards, and benchmarks
 - Operationalized at each grade-level through grade-level performance standards (GLPS's)
- Alternate Assessment Framework
 - Organized by the same strands, standards, and benchmarks
 - Operationalized through Expanded Grade Band Expectations (EGBE's) aligned with grade-level performance standards
- EGBE's organized under four different communication levels in order to meaningfully include all students in system
- Achievement Level Descriptors

Participants – Description of Raters

Raters were recruited and enlisted by the New Mexico Public Education Department to represent subject matter expertise (literacy, science, mathematics, or special education) and to represent the state geographically. Twenty-six different teachers attended the alignment workshops across the three days. Some of the teachers attended more than one day depending on their breadth of knowledge within each of the respective content areas. Thirteen teachers participated in the Math workshop, 13 for Science and 15 attended the Language Arts workshop. Some overlap of special education teachers was apparent; five teachers attended for all three days.

Demographics	Mathematics	Science	Language Arts
Total Teachers	13	13	15
Male	2	3	2
Female	11	10	13
Average Years Experience	16	16	15
General Ed Teachers	2	7	6
Special Education Teachers	11	6	9
Advanced Degree	8	9	10

Teachers had a wealth of grade-level experience, ranging from PreK-12th grade. We made a focused attempt to include a general education content teacher in each grade-level group, but more than half of the participants were Special Education teachers. The table below provides teacher demographics and includes the content area in which they were licensed.

Language Arts Teacher Demographics

Name	Degree	Years	Grades	Sped	Content
Adamson, Aileen	BA English	17	4-12	No	
Brady, Jon	BA k-8, MA k-12	10	4/5	Yes	English/science/social studies
Burtolino, Vicki	BA - English/Theatre arts, MA - Comm. Studies, endorsements TESOC & Literary science	35	6-12 +college	No	ESL 6-12, college freshmen English
Claussen, Sarah	BS in Ed, MST	14	1-4	No	
Dixon, Veronica	BS/Elem Ed 1-6, Sped k-12, T-sol end.	10	4/5	Yes	
Gossett, Kathy	BA - Sped, MA - Sped	10	9-12	Yes	
Holder, Melissa	BA - Reg Ed, MA - Reg Ed	10	6	No	
McKibben, Rawnie	BA, Sped English, MA, Sped	14	9-12	Yes	
McMahan, Kelly	BA, licenses in GE & Sped	17	6-8	Yes	
Ocampo, Kathleen	BA - SPED	4	4	Yes	
Ortega, DeNanna	MA Admin	21	K-12	Yes	
Riker-Tinguely, Claire	BA, MA, Ed Admin	12	6-9, 9-12	No	
Saavedra, Terri	BS/MST	9	7/4	No	
Valerma, Julia	M.Ed	16	K-12	Yes	
Vollrath, Ian	BS Sped k-12, M.Ed counseling and human relations	27	K-12	Yes	English

Every teacher participating in the alignment of language arts had a bachelor's or Master's degree (n=15). The degrees spanned in specificity, many in Education, Special Education, and Education Administration (n=12). They averaged 15 years of experience. Grades taught ranged from k-12, with most teaching 4th grade and higher (n=11). More than half of the teachers were involved in teaching Special Education.

Math Teacher Demographics

Name	Degree	Years	Grades	Sped (yrs)	Current	Subject
Saacedra, Terri	BS/MST	9	7/4	6 Inclusion	4th	All
Claussen, Sarah	AA/BS in Ed/MST	14	1, 2, 3, 4	Inclusion	4th	Inclusion
Lopez, Pat	MA Sped	17	K-12	Inclusion /self contained	Director	Director
Dixon, Veronica	Sped	10	4-5	low income (10), Inclusion, self contained	4/5	Experience GMD
Addamson, Aileen	AA, BA English	16	4, 5, 6	Inclusion 6 yrs.	4th	
Guame, Mike	Ba Math, Ma EmD	33	9-12	0	High school principal	
Ezell, Greag	BS/MST	26	7-12	0	HS Math/Science	
Trujillo, Lillian	BS Sped	11	1-5, 9-12	Self contained, Inclusion	10-12	Inclusion
Bustos, Mary	MA C&I	20	9-12	Inclusion, 2 yrs.	9-12	General ed and Inclusion
McKibben, Rawnie	BS SpEd/MA Sped	14	Head start - 12	Sped	9-12	Sped, self contained
Gossett, Kathy	BS Sped, MA Sped	10	9-12	Sped	9-12	Sped, self contained
McMahan, Kelly	BA, GenEd & Sped	17	K-8	15	6-8	ISP - intensive, support - severe
Valrema, Julia	Med, SpEd	16	K-8	16	7-8	Resource room and w/ teacher

Most teachers participating in the alignment of mathematics had a master's degree (n=8) in Education or Special Education. They averaged 16 years of experience. Average grades taught ranged from K-12, many teachers having experience in multiple grades. Inclusion was the most common Special Education field representing half of the participating teachers (n=7). Current grades taught ranged from 4th grade through high school, with the middle and high school teachers teaching several grade bands. A variety of subjects were taught including general and special education, resource room, and intensive to severe ISP support.

Science Teacher Demographics

Name	Degree	Years	Content	Sped	Grades Taught
Dixon, Veronica	BA, k-12 Sped, T-SOL, 1-6 Elem Ed	10	High school, Elem and middle school	10 yrs. Sped	
North, Cathe	BA, k-12 Elem, MS, k-12 Sped, EDS, k-12 Admin			10 yrs	
Adamson, Aileen	BA English	17			4-12th grade
McKibben, Rawnie	BA - Sped, MA Sped	14			All subjects, head start
Vollrath, Ian	BS - Sped k-12, M.Ed - counseling human relations	27			K-12 (all)
Hawf, Stephen	BA Elem, Sped, M.Ed reading	21	K-12 Music		all, lately HS science
Claussen, Sarah	AA, BS in Ed, MST	14		Both (Inclusion)	1, 2, 3, 4
Saavedra, Terri	BS k12/MST	9		Inclusion	7, 4
Bossett, Kathy	BA, Sped k-12, MA, Sped	10		Self contained all subjects	9-12
Valerma, Julia	M.Ed, Sped	16		MS resource room	k-12
Tinguely, Robert	BS Biology, M.Ed C&I	14	Science		9-12
Roney, Sandra	BA Political Sci/Sped	13	Gen Ed, Sped, Voc Rehab		9-12
Davis, Jeanette	Ba Elem Ed, MA Sp Ed	36		14 yrs	Pre-school thru grade school

Most teachers participating in the alignment of science had a master's degree (n=9). Most degrees specialized in education and special education. They averaged 16 years of experience. Teachers taught a variety of content including music, science, general and special education, and vocational rehabilitation. Half of the teachers had experience teaching special education (n=7) with specializations in inclusive settings, self-contained settings, and resource rooms. Grades taught ranged from pre-school-12th grade, with 4 teachers primarily teaching high school, and 2 primarily teaching in the pre-school and elementary grades.

Process: Training and Reaching Consensus

Teachers were trained by Drs. Tindal and Crawford in the morning and then asked to do the following for the rest of the day:

1. Rate the cognitive complexity of the Expanded Grade Band Expectations (EGBE's)
2. Link EGBE's to Grade-Level Performance Standards (GLPS's)
3. Rate the cognitive complexity of the alternate assessment items
4. Align alternate assessment items to EGBE's
5. Align Achievement Level Descriptors (ALDs) to EGBE's

They worked in grade-level groups with researchers serving as advisors. The grade level groups consisted of grade bands 3-4, 5-6, 7-8, 9-10, and 11-12. Each day began at 8:00 a.m. and concluded at 4:00 p.m. with a one-hour lunch break. Each day began with training on critical concepts associated with content standards and test item alignment. Teachers were trained on how to evaluate:

- Strength of linkage of EGBE's (Expanded Grade Band Expectations) to grade-level performance standards
- Depth of knowledge (or cognitive challenge) inherent in the EGBE's
- Depth of knowledge (or cognitive challenge) of the items on the alternate assessment
- Degree of alignment of assessment items to the EGBE's
- Strength of alignment between Achievement Level Descriptors and EGBE's

After the training, teachers were provided with the following materials: (1) a form to record their ratings, (2) the General Education Assessment Framework, (3) the Alternate Assessment Framework, and (4) the Achievement Level Descriptors provided by the New Mexico Public Education Department for each content area (Math, Science, and Language Arts).

Participants worked independently to rate the depth of knowledge (DOK) of the EGBE's on a scale of 1-4. When done judging DOK, they worked with their team members to come to a consensus score for each EGBE. The raters then used a 4-point scale to rate the linkage between each EGBE and each Grade-Level Performance Standard (GLPS). Finally, they aligned the

alternate assessment items with the EGBE's on categorical concurrence (and we calculated range of knowledge, balance of representation, and depth of knowledge).

After lunch, participants independently rated depth of knowledge of the alternate assessment items on the NMAPA (New Mexico Alternate Performance Assessment) for each content area. Then, they met to reach consensus. Participants then aligned these test items to the EGBE's. The day concluded with raters aligning the EGBE's to the Achievement Level Descriptors using a 3-point scale. Each evaluation was performed individually and then a group consensus was reached for each grade band group. Each day of the study proceeded in the manner described above, but with a focus on a different content area.

Alignment Workshop – NM Alternate Performance Assessment – Agenda 3/28-4/2

8:00 – 8:30	Continental breakfast and introductions
8:30 – 9:30	DOK training <ul style="list-style-type: none"> • Overheads and examples • Group work through • Check out
9:30 – 11:00	DOK analysis of EGBE's Break as needed
11:00 – 12:00	Linkage of EGBE's to GLPS <ul style="list-style-type: none"> • Training • Confirmatory analysis
12:00 – 12:45	Lunch and update
12:45 – 1:30	DOK analysis of assessment items
1:30 – 3:00	Alignment of test items to EGBE's <ul style="list-style-type: none"> • Training • Analysis Break as needed
3:00 – 3:30	Achievement level descriptors training
3:30 – 4:00	Achievement level descriptors (ALD's) to EGBE's <ul style="list-style-type: none"> • Analysis Level 1: Is each descriptor aligned with • Analysis Level 2: List specific EGBE's aligned with each ALD

Depth of Knowledge Rating Values and Definitions: EGBE's and Items

1	Recognition and reproduction: Recognition or reproduction of a fact, information, or procedure	<i>Attend:</i> touch, look, vocalize, respond, recognize
2	Skill and concept: Use information or conceptual knowledge (may require two or more steps)	<i>Memorize or Recall:</i> list, describe, identify, state, define, label
3	Strategic thinking: Requires reasoning, developing a plan or a sequence of steps, some complexity, more than one possible answer (non-routine problem-solving)	<i>Comprehend:</i> explain, conclude, group, restate, review, translate
4	Extended thinking: Requires an investigation, time to think and process multiple conditions of the problem (e.g., completing a project, including how to design and execute it)	<i>Apply:</i> compute, organize, collect, classify, construct, solve, operate, use, generalize

Linkage Rating Values and Definitions: EGBEs to GLPS's

1	<i>No linkage with grade level benchmarks</i>
2	<i>Vaguely linked with substantial reduction in breadth and depth</i>
3	<i>Clearly linked with substantial reduction in breadth and/or depth</i>
4	<i>Aligned (perfect match) with no reduction in breadth and depth</i>

Alignment Rating Values and Definitions: Items to EGBEs

1	<i>Not aligned</i>
2	<i>Somewhat aligned</i>
3	<i>Clearly aligned</i>

Section III – Analysis of Extended Grade Band Expectations (EGBE's)

Each EGBE was analyzed for linkage to the Grade Level Performance Standards (GLPS) and for the Depth of Knowledge (DOK) on separate 4-point scales. For each grade band, the results are expressed as a frequency count and as a percentage. Note: Each EGBE was designed to fit more than one GLPS so the number of EGBE's rated on depth of knowledge (DOK) is lower than the number of EGBE's linked to Grade Level Performance Standards.

We analyzed the results from individual judges and found that they varied in harshness/leniency resulting in varied correlations among judgments. This finding was not surprising given that the training was general and needed to be specifically applied to the content of the EGBEs and GLPS's of various grade levels. We required judges to make individual ratings and then reach a common rating. As we observed their discussion, this step actually had the effect of strengthening the certainty of final judgments made by each group with variation discussed out.

As would be expected, and by design, the vast majority of EGBE's were linked to the GLPS's quite well (in the range of 2-3) and were generally in the range of 2-3 on the DOK. The EGBE's are expected to reflect a substantial reduction of breadth and complexity (2-3) so this linkage level is not surprising. Furthermore, the depth of knowledge (DOK) is not expected to reflect 'extended thinking' (4) and should include more than rote recall (1); therefore, the range of 2-3 is to be expected. Basically, the EGBE's reflect a broad range of linkages and difficulties to the GLPS's, thereby reflecting a range of reduction in the breadth and complexity. Occasionally, in some grades and content areas, the number and percentage of 1's or 4's reaches a substantial number (in the 15%-25% range) but this is fairly infrequent.

In all of the tables below and for each subject area and grade band, the number of EGBE's and then the percentage of EGBE's are reported. *See Appendix A1 (language arts), A2 (mathematics), and A3 (science)*

Language Arts: Depth of Knowledge and EGBEs to Grade Level Performance Standards

Grade 3-4

<i>Scale</i>	<i>DOK</i>
Recall	10
Skill-Concept	17
Strategic	13
Extended	12
Total	52

<i>Scale</i>	<i>Linkage</i>
No Link	4
Vague Link	45
Clear Link	74
Aligned	33
Total	156

<i>Scale</i>	<i>DOK</i>
Recall	19.2%
Skill-Concept	32.7%
Strategic	25.0%
Extended	23.1%
Total	100.0%

<i>Scale</i>	<i>Linkage</i>
No Link	2.6%
Vague Link	28.8%
Clear Link	47.4%
Aligned	21.2%
Total	100.0%

Grade 5-6

<i>Scale</i>	<i>DOK</i>
Recall	0
Skill-Concept	13
Strategic	15
Extended	4
Total	28

<i>Scale</i>	<i>Linkage</i>
No Link	4
Vague Link	24
Clear Link	45
Aligned	1
Total	73

<i>Scale</i>	<i>DOK</i>
Recall	0.0%
Skill-Concept	46.4%
Strategic	53.6%
Extended	14.3%
Total	100.0%

<i>Scale</i>	<i>Linkage</i>
No Link	5.5%
Vague Link	32.9%
Clear Link	61.6%
Aligned	1.4%
Total	100.0%

Grade 7-8

<i>Scale</i>	<i>DOK</i>
Recall	7
Skill-Concept	8
Strategic	8
Extended	3
Total	26

<i>Scale</i>	<i>Linkage</i>
No Link	1
Vague Link	14
Clear Link	48
Aligned	0
Total	63

<i>Scale</i>	<i>DOK</i>
Recall	26.9%
Skill-Concept	30.8%
Strategic	30.8%
Extended	11.5%
Total	100.0%

<i>Scale</i>	<i>Linkage</i>
No Link	1.6%
Vague Link	22.2%
Clear Link	76.2%
Aligned	0.0%
Total	100.0%

Grade 9-10

<i>Scale</i>	<i>DOK</i>
Recall	3
Skill-Concept	4
Strategic	13
Extended	13
Total	33

<i>Scale</i>	<i>Linkage</i>
No Link	18
Vague Link	27
Clear Link	22
Aligned	8
Total	75

<i>Scale</i>	<i>DOK</i>
Recall	9.1%
Skill-Concept	12.1%
Strategic	39.4%
Extended	39.4%
Total	100.0%

<i>Scale</i>	<i>Linkage</i>
No Link	24.0%
Vague Link	36.0%
Clear Link	29.3%
Aligned	10.7%
Total	100.0%

Grade 11-12

<i>Scale</i>	<i>DOK</i>
Recall	5
Skill-Concept	3
Strategic	18
Extended	10
Total	36

<i>Scale</i>	<i>Linkage</i>
No Link	9
Vague Link	17
Clear Link	26
Aligned	2
Total	54

<i>Scale</i>	<i>DOK</i>
Recall	13.9%
Skill-Concept	8.3%
Strategic	50.0%
Extended	27.8%
Total	100.0%

<i>Scale</i>	<i>Linkage</i>
No Link	16.7%
Vague Link	31.5%
Clear Link	48.1%
Aligned	3.7%
Total	100.0%

Mathematics: Depth of Knowledge and EGBEs to Grade Level Performance Standards

Grade 3-4

<i>Scale</i>	<i>DOK</i>
Recall	6
Skill-Concept	17
Strategic	13
Extended	4
Total	40

<i>Scale</i>	<i>Linkage</i>
No Link	13
Vague Link	43
Clear Link	52
Aligned	17
Total	125

<i>Scale</i>	<i>DOK</i>
Recall	15.0%
Skill-Concept	42.5%
Strategic	32.5%
Extended	10.0%
Total	100.0%

<i>Scale</i>	<i>Linkage</i>
No Link	10.4%
Vague Link	34.4%
Clear Link	41.6%
Aligned	13.6%
Total	100.0%

Grade 5-6

<i>Scale</i>	<i>DOK</i>
Recall	3
Skill-Concept	11
Strategic	8
Extended	4
Total	26

<i>Scale</i>	<i>Linkage</i>
No Link	6
Vague Link	20
Clear Link	6
Aligned	5
Total	37

<i>Scale</i>	<i>DOK</i>
Recall	11.5%
Skill-Concept	42.3%
Strategic	30.8%
Extended	15.4%
Total	100.0%

<i>Scale</i>	<i>Linkage</i>
No Link	16.2%
Vague Link	54.1%
Clear Link	16.2%
Aligned	13.5%
Total	100.0%

Grade 7-8

<i>Scale</i>	<i>DOK</i>
Recall	15
Skill-Concept	7
Strategic	5
Extended	4
Total	31

<i>Scale</i>	<i>Linkage</i>
No Link	0
Vague Link	32
Clear Link	19
Aligned	1
Total	52

<i>Scale</i>	<i>DOK</i>
Recall	48.4%
Skill-Concept	22.6%
Strategic	16.1%
Extended	12.9%
Total	100.0%

<i>Scale</i>	<i>Linkage</i>
No Link	0.0%
Vague Link	61.5%
Clear Link	36.5%
Aligned	1.9%
Total	100.0%

Grade 9-12

<i>Scale</i>	<i>DOK</i>
Recall	7
Skill-Concept	12
Strategic	4
Extended	8
Total	31

<i>Scale</i>	<i>Linkage</i>
No Link	11
Vague Link	28
Clear Link	8
Aligned	0
Total	47

<i>Scale</i>	<i>DOK</i>
Recall	22.6%
Skill-Concept	38.7%
Strategic	12.9%
Extended	25.8%
Total	100.0%

<i>Scale</i>	<i>Linkage</i>
No Link	23.4%
Vague Link	59.6%
Clear Link	17.0%
Aligned	0.0%
Total	100.0%

Science: Depth of Knowledge and EGBEs to Grade Level Performance Standards

Grade 3-4

<i>Scale</i>	<i>DOK</i>
Recall	8
Skill-Concept	17
Strategic	14
Extended	7
Total	46

<i>Scale</i>	<i>Linkage</i>
No Link	5
Vague Link	47
Clear Link	66
Aligned	34
Total	152

<i>Scale</i>	<i>DOK</i>
Recall	17.4%
Skill-Concept	37.0%
Strategic	30.4%
Extended	15.2%
Total	100.0%

<i>Scale</i>	<i>Linkage</i>
No Link	3.3%
Vague Link	30.9%
Clear Link	43.4%
Aligned	22.4%
Total	100.0%

Grade 5-6

<i>Scale</i>	<i>DOK</i>
Recall	5
Skill-Concept	4
Strategic	6
Extended	2
Total	17

<i>Scale</i>	<i>Linkage</i>
No Link	2
Vague Link	14
Clear Link	16
Aligned	2
Total	34

<i>Scale</i>	<i>DOK</i>
Recall	29.4%
Skill-Concept	23.5%
Strategic	35.3%
Extended	11.8%
Total	100.0%

<i>Scale</i>	<i>Linkage</i>
No Link	5.9%
Vague Link	41.2%
Clear Link	47.1%
Aligned	5.9%
Total	100.0%

Grade 7-8

<i>Scale</i>	<i>DOK</i>
Recall	4
Skill-Concept	4
Strategic	9
Extended	4
Total	21

<i>Scale</i>	<i>Linkage</i>
No Link	4
Vague Link	10
Clear Link	9
Aligned	9
Total	32

<i>Scale</i>	<i>DOK</i>
Recall	19.0%
Skill-Concept	19.0%
Strategic	42.9%
Extended	19.0%
Total	100.0%

<i>Scale</i>	<i>Linkage</i>
No Link	12.5%
Vague Link	31.3%
Clear Link	28.1%
Aligned	28.1%
Total	100.0%

Grade 9-12

<i>Scale</i>	<i>DOK</i>
Recall	6
Skill-Concept	2
Strategic	9
Extended	4
Total	21

<i>Scale</i>	<i>Linkage</i>
No Link	0
Vague Link	20
Clear Link	9
Aligned	3
Total	32

<i>Scale</i>	<i>DOK</i>
Recall	28.6%
Skill-Concept	9.5%
Strategic	42.9%
Extended	19.0%
Total	100.0%

<i>Scale</i>	<i>Linkage</i>
No Link	0.0%
Vague Link	62.5%
Clear Link	28.1%
Aligned	9.4%
Total	100.0%

Section IV – Item Analysis of the New Mexico Alternate Performance Assessment

Each item in the New Mexico Alternate Performance Assessment was coded for Depth of Knowledge (DOK) and for alignment with a primary Extended Grade Band Expectation (EGBE). Following are the results from this analysis using two scales (see Section II-Methods). In general, the items ranged in DOK with the vast majority in the 2-3 range, reflecting a well-balanced blend of skills and concepts (2) along with strategic thinking (3); the number of items rated 1 (rote recall) and 4 (extended thinking) was less, particularly those on the high end (rated 4). The alignment with the EGBEs was generally quite strong with only a few items rated 1 (not well aligned). This finding was true for all subject areas and all grade bands.

In all of the tables below and for each subject area and grade band, the number of items and then the percentage of items are reported. *See Appendix B1 (language arts), B2 (mathematics), and B3 (science).*

Language Arts

Grade 3-4

<i>Scale</i>	<i>DOK</i>
Recall	13
Skill-Concept	37
Strategic	11
Extended	0
Total	61

<i>Scale</i>	<i>Aligned</i>
Not	0
Somewhat	0
Clearly	61
Total	61

<i>Scale</i>	<i>DOK</i>
Recall	21%
Skill-Concept	61%
Strategic	18%
Extended	0%
Total	100%

<i>Scale</i>	<i>Aligned</i>
Not	0%
Somewhat	0%
Clearly	100%
Total	100%

Grade 5-6

<i>Scale</i>	<i>DOK</i>
Recall	22
Skill-Concept	29
Strategic	8
Extended	0
Total	59

<i>Scale</i>	<i>Aligned</i>
Not	0
Somewhat	17
Clearly	42
Total	59

<i>Scale</i>	<i>DOK</i>
Recall	37%
Skill-Concept	49%
Strategic	14%
Extended	0%
Total	100%

<i>Scale</i>	<i>Aligned</i>
Not	0%
Somewhat	29%
Clearly	71%
Total	100%

Grade 7-8

<i>Scale</i>	<i>DOK</i>
Recall	27
Skill-Concept	23
Strategic	6
Extended	0
Total	56

<i>Scale</i>	<i>Aligned</i>
Not	0
Somewhat	4
Clearly	52
Total	56

<i>Scale</i>	<i>DOK</i>
Recall	48%
Skill-Concept	41%
Strategic	11%
Extended	0%
Total	100%

<i>Scale</i>	<i>Aligned</i>
Not	0%
Somewhat	7%
Clearly	93%
Total	100%

Grade 9-10

<i>Scale</i>	<i>DOK</i>
Recall	9
Skill-Concept	20
Strategic	23
Extended	0
Total	52

<i>Scale</i>	<i>Aligned</i>
Not	4
Somewhat	35
Clearly	13
Total	52

<i>Scale</i>	<i>DOK</i>
Recall	17%
Skill-Concept	38%
Strategic	44%
Extended	0%
Total	100%

<i>Scale</i>	<i>Aligned</i>
Not	8%
Somewhat	67%
Clearly	25%
Total	100%

Grade 11-12

<i>Scale</i>	<i>DOK</i>
Recall	10
Skill-Concept	19
Strategic	19
Extended	4
Total	52

<i>Scale</i>	<i>Aligned</i>
Not	8
Somewhat	16
Clearly	24
Total	48

<i>Scale</i>	<i>DOK</i>
Recall	19%
Skill-Concept	37%
Strategic	37%
Extended	8%
Total	100%

<i>Scale</i>	<i>Aligned</i>
Not	17%
Somewhat	33%
Clearly	50%
	100%

Mathematics

Grade 3-4

<i>Scale</i>	<i>DOK</i>
Recall	8
Skill-Concept	27
Strategic	20
Extended	1
Total	56

<i>Scale</i>	<i>Aligned</i>
Not	0
Somewhat	13
Clearly	40
Total	53

<i>Scale</i>	<i>DOK</i>
Recall	14%
Skill-Concept	48%
Strategic	36%
Extended	2%
Total	100%

<i>Scale</i>	<i>Aligned</i>
Not	0%
Somewhat	25%
Clearly	75%
Total	100%

Grade 5-6

<i>Scale</i>	<i>DOK</i>
Recall	1
Skill-Concept	46
Strategic	9
Extended	0
Total	56

<i>Scale</i>	<i>Aligned</i>
Not	5
Somewhat	6
Clearly	43
Total	54

<i>Rating Scale</i>	<i>DOK</i>
Recall	2%
Skill-Concept	82%
Strategic	16%
Extended	0%
Total	100%

<i>Scale</i>	<i>Aligned</i>
Not	9%
Somewhat	11%
Clearly	80%
Total	100%

Grade 7-8

<i>Scale</i>	<i>DOK</i>
Recall	23
Skill-Concept	20
Strategic	14
Extended	0
Total	57

<i>Scale</i>	<i>Aligned</i>
Not	1
Somewhat	12
Clearly	44
Total	57

<i>Scale</i>	<i>DOK</i>
Recall	40%
Skill-Concept	35%
Strategic	25%
Extended	0%
Total	100%

<i>Scale</i>	<i>Aligned</i>
Not	2%
Somewhat	21%
Clearly	77%
Total	100%

Grade 9-12

<i>Scale</i>	<i>DOK</i>
Recall	6
Skill-Concept	29
Strategic	13
Extended	3
Total	51

<i>Scale</i>	<i>Aligned</i>
Not	2
Somewhat	6
Clearly	43
Total	51

<i>Scale</i>	<i>DOK</i>
Recall	12%
Skill-Concept	57%
Strategic	25%
Extended	6%
Total	100%

<i>Scale</i>	<i>Aligned</i>
Not	4%
Somewhat	12%
Clearly	84%
Total	100%

Science

Grade 3-4

<i>Scale</i>	<i>DOK</i>
Recall	7
Skill-Concept	38
Strategic	13
Extended	2
Total	60

<i>Scale</i>	<i>Aligned</i>
Not	0
Somewhat	10
Clearly	45
Total	55

<i>Scale</i>	<i>DOK</i>
Recall	12%
Skill-Concept	63%
Strategic	22%
Extended	3%
Total	100%

<i>Scale</i>	<i>Aligned</i>
Not	0%
Somewhat	18%
Clearly	82%
Total	100%

Grade 5-6

<i>Scale</i>	<i>DOK</i>
Recall	6
Skill-Concept	17
Strategic	29
Extended	6
Total	58

<i>Scale</i>	<i>Aligned</i>
Not	0
Somewhat	6
Clearly	53
Total	59

<i>Scale</i>	<i>DOK</i>
Recall	10%
Skill-Concept	29%
Strategic	50%
Extended	10%
Total	100%

<i>Scale</i>	<i>Aligned</i>
Not	0%
Somewhat	10%
Clearly	90%
Total	100%

Grade 7-8

<i>Scale</i>	<i>DOK</i>
Recall	12
Skill-Concept	22
Strategic	24
Extended	3
Total	61

<i>Scale</i>	<i>Aligned</i>
Not	0
Somewhat	7
Clearly	54
Total	61

<i>Scale</i>	<i>DOK</i>
Recall	20%
Skill-Concept	36%
Strategic	39%
Extended	5%
Total	100%

<i>Scale</i>	<i>Aligned</i>
Not	0%
Somewhat	11%
Clearly	89%
Total	100%

Grade 9-12

<i>Scale</i>	<i>DOK</i>
Recall	4
Skill-Concept	19
Strategic	26
Extended	6
Total	55

<i>Scale</i>	<i>Aligned</i>
Not	2
Somewhat	12
Clearly	32
Total	46

<i>Scale</i>	<i>DOK</i>
Recall	7%
Skill-Concept	35%
Strategic	47%
Extended	11%
Total	100%

<i>Scale</i>	<i>Aligned</i>
Not	4%
Somewhat	26%
Clearly	70%
Total	100%

Section V – Summary Statistics of Alignment using Webb’s Dimensions

Tindal² describes the dimensions from Webb’s alignment³ as follows:

1. *Categorical Concurrence*, or the degree to which standards and assessments address the same content categories. This criterion is met if both documents display the same or consistent categories of content.

2. *Depth-of-Knowledge Consistency*, or the degree to which the depth or complexity of knowledge required by the standards and assessments are in agreement. If the assessment is as demanding conceptually as the expectations standards set for the students, this criterion is met. Depth-of-knowledge is judged at four levels: (a) recall of fact, information, or procedure; (b) skill in using information, conceptual knowledge, or procedures of two or more steps; (c) strategic thinking, reasoning, developing a plan or sequence of steps, complexity, more than one possible answer, requiring less than 10 minutes to do; and, (d) extended thinking, requiring an investigation, time to think and process multiple conditions of the problem or task, and requiring more than 10 minutes to do non-routine manipulations.

3. *Range-of-Knowledge Correspondence*, or the degree to which the span of knowledge a standard expects of students matches that required to correctly answer the assessment items or activity.

4. *Balance of Representation*, or the extent to which assessment items are evenly dispersed across learning objectives within a standard.

The following results were obtained in reading, mathematics, and science across four grade bands: 3/4, 5/6, 7/8, and 9/12. In these tables, the grade band is listed in the left-most column followed by the number of items in the alternate assessment (N of items), the number of standards to which they were aligned (N of Stds), the subsequent percentage of this alignment of items to standards (% of Stds), the average range of items across all the specific standards-expectations in all standards (Ave. Range), Balance of Representation that depicts the proportion

² Tindal, G. (2006). Alignment of alternate assessments using the Webb system, in *Aligning Assessment to Guide the Learning of All Students*. Washington, D. C. Council of Chief State School Officers.

³ Webb, N. L. (2002). *Alignment study in language arts, mathematics, science, and social studies of state standards and assessments for four states*. Washington, DC: Council of Chief State School Officers (CCSSO): State Collaborative on Assessment and Student Standards (SCASS), Technical Issues in Large Scale Assessments (TILSA).

of items spread over the standards-expectations in which they were aligned (Bal. of Rep.), and then three analyses of Depth of Knowledge with the percentage of agreement in DOK between items and standards-expectations in which they were equal (% DOK Equal), in which the standard-expectation DOK exceeded the item's DOK (% DOK Std > AA) or visa versa with the standard-expectation DOK actually less than the item DOK (% DOK Std < AA).

In these four areas, the best results would be that all standards are addressed (categorical concurrence is 100%), that a substantial percentage of objectives are addressed (range is 25% to 65%), that objectives are evenly addressed (balance is .20 to .50), and a large percentage of items are at or slightly below the EGBE's on depth of knowledge. These expectations, however, are somewhat speculative given the brief history of assessment with this population. For the most part, the New Mexico Alternate Performance Assessment is within these ranges.

Language Arts All Grades: The findings were quite consistent across the grade bands.

- All of the standards reflected in the EGBE's were addressed with items in the assessment.
- Often, however, an EGBE had only a few items; a few had many items; many had no items in the assessment that addressed them. The range of knowledge, therefore, was low to moderate with as few as 15% of the EGBE's covered to a maximum of 23%.
- Similarly, balance of representation was also low to moderate, ranging from .28 to .46.
- Finally, as expected, depth of knowledge (DOK) was either equal between the EGBE's and items or the EGBE's exceeded the items in DOK.

Grade Band	N of Items	N of Stds	% Stds	Ave. Range	Bal. of Rep.	% DOK Std = AA	% DOK Std < AA	% DOK Std > AA
3/4	61	3	100	20%	.28	43	28	30
5/6	59	3	100	23%	.38	17	0	83
7/8	56	2	100	15%	.46	48	11	41
9/10	52	2	100	15%	.45	36	17	48
11/12	50	3	100	23%	.34	26	20	54

Mathematics All Grades: The findings on all dimensions were quite consistent across the grades.

- All of the extended grade band expectations (EGBE's) were typically addressed with at least one item. Typically, each standard or strand had several items.
- Most of the specific objectives (EGBE's) had items that were in the Alternate Assessment. The range of knowledge reflected across EGBE's within standards was moderate, above 30% and only once at 23%.
- The balance of representation, reflecting the distribution of items within standards, was quite even in all but grade band 3-4 (which was low): Most objectives had more than one item while few objectives had an inordinate number of objectives.
- Finally, depth of knowledge (DOK) was judged for both the EGBE's and the items as expected. DOK was higher in the EGBE's than in the items. This outcome was expected, given the need to reduce the complexity of alternate assessments for this population. Generally, the percentages were quite evenly split between standards that were equal or above the DOK for items.

Grade Band	N of Items	N of Stds	% Stds	Ave. Range	Bal. of Rep.	% DOK Std = AA	% DOK Std < AA	% DOK Std > AA
3/4	56	5	100	31%	.25	41	9	50
5/6	54	5	100	38%	.46	63	6	31
7/8	56	4	100	31%	.39	41	2	57
9/12	47	3	100	23%	.44	47	11	43

Science All Grades. As in both reading and Mathematics, science was consistent across grades.

- All standards reflected in the EGBE's were addressed with at least one item and usually several items.
- Range of specific expectations hit within a standard, however, varied from 15% (in two grade bands of 5/6 and 7/8) to a high of 23% in two grade bands (3/4 and 9/12).
- Balance of representation was very high. This outcome means that an even proportion of items are spread across the EGBE's that were addressed.
- The EGBE's and items were quite even in the depth of knowledge with the majority usually equal but a substantial percentage above or below; only grade 9/12 had an uneven percentage with very few items rated with a higher DOK than the EGBE.

Grade Band	N of Items	N of Stds	% Stds	Ave. Range	Bal. of Rep.	% DOK Std = AA	% DOK Std < AA	% DOK Std > AA
3/4	55	3	100	23%	.65	38	24	38
5/6	59	3	67	15%	.58	44	25	31
7/8	56	2	100	15%	.62	52	16	31
9/12	46	3	100	23%	.64	48	9	43

See Appendix C (language arts, mathematics and science)

Section VI – Alignment of Achievement Level Descriptor to EGBE’s

Using a general 3-point scale of no alignment, some alignment, and clear alignment (as was used in rating the items to EGBE’s), we documented the relation between the Achievement Level Descriptors (ALD’s) and the EGBE’s. This was the last activity of each day. Teachers were given a 1-2 page descriptor of the ALDs and asked to judge how well they reflected or represented the content of the EGBE for that subject area and grade band.

Following are the results from this analysis. As can be seen in the tables for each subject area, the majority of the ALDs appeared to be consistent or aligned with the content of the EGBE. Only occasionally are ratings of no relation or only some relation present (five times in language arts, none in writing, twice in mathematics, and ten times in science).

Language Arts	Advanced	Proficient	Nearing Proficiency	Beginning Step
3 to 4	3	3	3	3
5 to 6	3	3	3	3
7 to 8	3	3	3	3
9 to 10	2	2	2	3
11 to 12	1 to 2	2	3	3

Writing	Advanced	Proficient	Nearing Proficiency	Beginning Step
3 to 4	3	3	3	3
5 to 6	3	3	3	3
7 to 8	3	3	3	3
9 to 10	3	3	3	3
11 to 12	3	3	3	3

Mathematics	Advanced	Proficient	Nearing Proficiency	Beginning Step
3 to 4	3	3	3	3
5 to 6	3	3	3	3
7 to 8	3	3	3	3
9 to 10	3	2	2	3
11 to 12	3	3	3	3

Science	Advanced	Proficient	Nearing Proficiency	Beginning Step
3 to 4	2	2	1	3
5 to 6	3	3	3	3
7 to 8	2	3	2	3
9 to 10	3	2	2	3
11 to 12	2	2	3	2

Section VII – Conclusions and Recommendations

The analysis of EGBE linked to GLPS revealed some grade bands in which the DOK and/or the linkage were high, perhaps too high.

1. *Language arts.* For grade bands 3-4, 9-10, and 11-12, about 23% to 40% of the EGBE's were rated 4) (extended thinking) on DOK, which may be too high; for the linkage in grade 3-4, nearly 1 in 5 EGBE was rated as not reduced for linkage, which may relate to the high level of DOK for that grade band.
2. *Mathematics.* Only grade band 9-12 had high ratings of DOK with over 1 in 4 EGBE rated 4 (extended thinking), which may be too high for that percentage.
3. *Science* had two grade bands with high DOK ratings (7-8 and 9-12) with nearly 1 in 5 rated 4 (extended thinking). For linkage, two grade bands had high ratings perhaps reflecting insufficient reduction of breadth with over 1 in 5 rated 4: Grade 3-4 and 7-8.

The analysis of assessment items for Depth of Knowledge and alignment to EGBE's revealed a few places where the New Mexico Department of Education may want to consider slight changes.

1. DOK was very low for a large percentage of items (equal to a 1 reflecting rote recall) in language arts for grades 7-8 (48% of the items) and for mathematics in grades 7-8 (40% were rated rote recall).

2. Webb alignment dimensions were low to moderate in range for *language arts*, which may need to be adjusted by the department. Basically, a wider variety of items, addressing more EGBE's, may need to be developed.
3. The DOK also may need to be adjusted in grade band 3-4 with 83% of the items rated below the EGBE DOK. This is somewhat in contrast to the earlier findings of the EGBE's in which they were rated as being high (relative to the GLPS).
4. The Webb analysis in *mathematics* revealed very even range of EGBE's covered (in terms of percentage) across all the grades; however, balance was quite low in grade band $\frac{3}{4}$. The Department may need to shift some of the coverage of EGBE's by redistributing items within those EGBE's that they cover for this grade band.
5. In *science*, the primary issue to be addressed is the low percentage of EGBE's addressed in grade bands $\frac{5}{6}$ and $\frac{7}{8}$; otherwise, the balance is high which means that, of those that are covered, the distribution is excellent. Finally, the DOK is quite high in a large percentage of items in all but the last grade band (9/12) with as many as 1 in 4 items above the DOK of the EGBE.