

Mapping cognitive pathways in mastering long division: A case study of grade 5-6 learners
supported with a dynamic model of proximal assessment and learner diagnosis

Madhabi Chatterji and Nancy Koh
Teachers College, Columbia University

Howard Everson
Fordham University

Pearl Solomon
St. Thomas Aquinas College

For reprints, please contact:

Madhabi Chatterji, Ph.D
Associate Professor of Measurement and Evaluation
Co-Director, Assessment and Evaluation Research Initiative (AERI)
Teachers College Columbia University
525, W 120th St.
New York, NY 10040.
E-mail: mb1434@columbia.edu

Paper presented at the annual meeting of the American Educational Research Association on

March 28, 2008 at New York, USA

Mapping cognitive pathways in mastering long division: A case study of grade 5-6 learners supported with a dynamic model of proximal assessment and learner diagnosis

Abstract

This paper presents a single-classroom case study of learners as they transitioned from grade 5-6, documenting their learning gaps, changes in cognitive development, and mastery of one mathematics domain, when supported with a *Proximal Assessment for Learner Diagnosis* (PALD) model by their teachers. PALD is a dynamic approach to classroom assessment that draws on the combined literatures in the cognitive and assessment sciences. In this study, we focus on learners at three developmental levels transacting with situated problems in the Long Division (LD) domain, a formal part of the mathematics curriculum at the school district research site. The paper first documents the operational components of the PALD model, showing examples of PALD practices used by teachers who taught the class in grades 5-6, and provides the foundational theory. It then combines quantitative analyses and qualitative expert reviews of students' work to identify different ways in which children at different developmental levels "think" as they gain expertise in the LD domain. While the class as a whole made visible gains in one year, those who were developmentally most delayed initially were left behind. The case analysis shows specifically where learning stalled or showed spurts, and how the PALD approach can effectively support student learning, highlighting implications for math pedagogy, diagnostic assessment, and teachers' assessment capacities.

Mapping cognitive pathways in mastering long division: A case study of grade 5-6 learners supported with a dynamic model of proximal assessment and learner diagnosis

Educational assessments that provide a close-up view of how children learn and develop in a domain, incorporating up-to-date knowledge in the cognitive and learning sciences, are still rare or absent in both classroom and large scale applications in the U.S today. A relatively recent national academy report called for more research in this particular area, emphasizing that design efforts directly engage teachers (Pellegrino, Chudowski, & Glaser, 2001). Coupled with that call, for some time now, there has also been a cry for more student-centered and formatively-oriented classroom assessment (Shepard, 2000; Stiggins, 2002; Wiggins, 1998).

The notion of formative classroom assessment, linked to instruction, and oriented towards developing learner proficiency in defined areas, is not a new concept (see Bloom, 1982; Gagne, 1997; Nitko, 1989). Over time, the spirit of formative assessment been evidenced in both general and special education environments under various protocols and labels (see for examples, Fuchs & Fuchs, 2003; Guskey, 2003). The idea has been revived today as a counter-movement to high stakes testing accompanying the national push for educational excellence and equitable outcomes under the NCLB Act of 2001. In recent applications, however, formative uses of assessment by classroom teachers have relied too often on data from state-administered, standards-based tests (on which schools are held accountable). Alternatively, such practices have depended on locally-developed tests that mimic state instruments in terms of item content and structure.

Externally-designed, standards-based tests rarely yield the fine-grained, diagnostic information necessary for teachers to effectively detect learner gaps along developmental

continua as students grapple with and gain expertise in particular domains. Although higher student performance might be evidenced on results of external tests as a consequence of the popular practice of “teaching to the test”, such spikes in the data are temporary, and student gains are rarely sustained when cognitive gaps remain unaltered.

There is thus, a high need today for research that focuses on developing models of classroom assessment that are *learning-centered*, and that can concurrently develop in teachers, the necessary assessment skills and attitudes to assess learners proximally and diagnostically, close to where the teaching and learning processes unfold in classrooms every day. The aim of reported research study, which forms one part of a larger research and development effort supported by the National Science Foundation, is to address these very needs.

Purpose

This paper reports a descriptive case study showing how one class of elementary students responded to a dynamic approach to classroom assessment titled, *Proximal Assessment for Learner Diagnosis* (PALD), and how their learning gaps were addressed over a two year period. The PALD approach, elaborated next, draws on the current and combined knowledge base in the cognitive, learning and assessment sciences. The case study follows the same group of learners over a two-year period as they transitioned from grade 5-6, documenting their cognitive development and thinking in one mathematics domain, when supported with a Proximal Assessment for Learner Diagnosis (PALD) model by their teachers. Specifically, the study focuses on learners at three developmental levels transacting with situated problems in the Long Division (LD) domain, a formal part of the mathematics curriculum at the school district research site.

The larger research project on the PALD approach began 3 years ago (Chatterji & Gordon, 2004), with field studies occurring between September 2005-June, 2007. Because teachers hold the key to sound assessment implementation in classrooms, the PALD model was formally field-tested and refined in partnership with an upstate New York school district, and with the involvement of real teachers and students. Four elementary schools and 14 classroom teachers were engaged as “PALD participants” in the 2-year long intervention development and formative research project, along with 30 classrooms and teachers participating as “ Non-PALD” participants for comparison (Total $N_{\text{students}}=700-800$; Total $N_{\text{teachers}}=44$). To ensure that the selected mathematics domains for the research project had local relevance, we drew on mathematics content standards provided by the state of New York, as adopted by the district in the form of curriculum maps. These maps served as guiding frameworks for designing project-related assessments for classroom and end-of-project use.

It is important to clarify at the start that the PALD research project incorporated a teacher development component, focusing on changing teachers’ assessment attitudes and practices; that component is not the focus of this paper. Rather, the present paper deals with learners’ cognitive processes, learning gaps, and outcomes when exposed to PALD practices, using case study methodology. Specific objectives of the present paper are as follows:

- 1) To describe the PALD model, present specific operational examples of how the PALD model plays out in actual classrooms, and discuss the theoretical foundations of the same.
- 2) To document, using three different methods/perspectives, how thinking and expertise in long division changed in one selected classroom of learners, as evidenced in their responses to specially-designed, *developmentally-diagnostic* assessments administered at

the end of grade 5 and grade 6. (Developmentally-diagnostic assessments are specific to the PALD approach).

- 3) To make inferences from the case data about learning and cognitive needs of learners at three developmental levels as they transitioned through grades 5-6.

The case analysis shows how the PALD approach can effectively support student learning, highlighting implications for math pedagogy, diagnostic assessment, and teachers' assessment capacities/skills.

Operational Elements of the *Proximal Assessment for Learner Diagnosis* (PALD) Model

PALD is characterized as a “dynamic” approach to classroom assessment. It calls on teachers to make cyclic and frequent use of informal and formal diagnostic assessments, so as to detect and address learning needs of students in selected domains, as they teach. According to the model, teachers' use of diagnostic assessment should be followed immediately with error analysis, mediation (taking corrective actions tailored to learner needs) and provision of intensive practice for learners to consolidate new learning and conceptual bonds.

Appendix A provides a diagrammatic representation of how the PALD model is expected to be implemented by a teacher, with a theoretical sequence of activities in planning, teaching, and concluding an instructional unit, marked as I-IV. The formative assessment and decision-making occurs during step III (PALD Implementation), and could involve as many cycles as deemed necessary and feasible by the teacher. Step III can be repeated after new material is introduced, with summative assessment (i.e., end-of-unit testing for assigning marks, making decisions on student progress and promotion to next level or unit) withheld until students' learning gaps are sufficiently addressed. Individual teacher judgment is required with regard to

the duration, continuity/discontinuity, and number of PALD cycles that can be embedded within a unit, based on the local curriculum scope and sequence, infrastructure, logistics, and schedules at different schools.

In an ideal implementation, the PALD approach is dynamic, because it could potentially involve adaptations and revisions to the plans for instruction, assessment, and to learning targets set for a unit, to suit needs of learners progressing at different paces and in different ways as they transact with a domain. This is indicated in the parts marked “REVISE” in the diagram shown.

The term “domain” has a specific connotation in the PALD approach. Instead of starting with a listing of instructional goals and objectives (or content area standards found in typical curriculum frameworks), the domain in the PALD approach emerges out of the “backwards analysis” from situated, integrative and real-life tasks that teachers would like students to master at the end of the lesson, unit, semester, or school year. Teachers connect a PALD domain with broader curriculum frameworks as a part of the long-term planning process marked under step I in Appendix A.

The focus of instruction and assessment planning in step II, centers around the type of culminating tasks in which the teachers would like students to gain mastery and domain of embedded concepts, skills, and competencies that flow out of the tasks. Box 1 below gives a concrete example of a culminating task at end-of-grade 5 level, from which the Long Division (LD) Domain was derived (focus of the present study).

Box 1. Culminating Task in Long Division Domain Used As Pre-Unit Assessment in a PALD Unit

Coconut cookies come in packets of 12. I would like to give 1 cookie to each 5th grader at (your school) to celebrate the last day of school (in June). There are 119 5th graders at (your school). How many packets of cookies should I buy so that every 5th grader gets at least 1 cookie? Will there be any cookies left over?

Show your plan to solve the problem. Then set up the numbers and solve it. Explain all the parts of your answer, and show how you would check if your answer is correct. Show all your work.

Instead of being guided by abstract statements, such as, *The student will apply a long division algorithm/ procedure taught and other operations needed to solve problems with 4-digit dividends and 2-digit divisors*, PALD teachers would select concrete, culminating tasks likely to be meaningful to learners as guides for designing of instruction and assessment. The targeted/valued (math) concept knowledge, vocabulary, cognitive skills, complex procedures and so on, are taught and tested through the medium of these “situated” real-life problems with which students are likely to make a connection. De-contextualized teaching and testing of relevant material (such as learning multiplication tables by themselves, outside a problem context), is not altogether excluded; rather, it is kept to a minimum and treated as supplementary to the PALD process (e.g., they can be used on the way to developing expertise in the culminating tasks that are identified as the key learning targets for students).

The backwards analysis from targeted tasks during domain specification, is expected to help teachers identify relevant types of prior knowledge that students need to attack the culminating problems and to be successful. In applying the PALD process with teachers, we asked them to envision and create such tasks, and to solve the problems first themselves so as to be able to identify all embedded concepts/skills via a concrete method. The backwards domain specification process is the means to help teachers identify the essential prior knowledge students need, to attack a new domain. It leaves the door open for teachers to test, gauge and address learners’ needs. By reinforcing, scaffolding, or teaching anew relevant but previously taught concepts that learners need to grasp *before* moving to new material in the domain, they foster learning development on a continuum (see PALD Implementation in step III in Appendix A).

As will become evident through concrete examples that follow in this section, the most useful assessments for PALD practices, are not only domain-referenced, but they also have

properties that reveal where children get stuck as they struggle to learn something new or different. They are labeled as “developmentally-diagnostic” assessments. Appendix B gives an example of a multi-part, developmentally-diagnostic assessment in the LD domain.

Developmentally-diagnostic PALD assessments comprise tasks that are ordered by complexity that tap backwards into students’ grasp of prior knowledge/skills relevant to gaining mastery of a relatively novel domain of performance. Assessment formats used by classroom teachers can be varied. They can be paper and pencil tests incorporating tasks such as those shown in Box 1. They can involve use of probing Q & A that teachers embed as they teach and build skills and concepts necessary for children to successfully solve the culminating performance tasks that serve as guiding targets. They can incorporate structured–response formats, or also include open-ended and performance assessments.

Regardless of format, to be diagnostically useful during teaching and learning, the task design, prompts and accompanying scoring mechanisms must elicit or clearly reveal students’ levels of understanding as well as misconceptions/ errors in a meaningful way to both teachers and students. An example of a student- and teacher-friendly analytic rubric used to encourage students to engage in self-monitoring their progress during a mini-unit on the LD domain follows in Box 2. Analytic scoring provides a simple way for tracking error patterns.

As may be evident, in theory, PALD involves creating a safe and nurturing assessment culture in the classroom, where learners are made aware of explicit learning targets. Affectively, the learners must feel good as they engage in learning. In theory, thus, they should not be afraid to closely examine their own learning gaps and errors, trusting their teachers to help them make gains, fill gaps, and achieve greater levels of expertise in a domain. Within practical limits,

PALD is meant to be implemented in an ongoing manner by teachers, with students and teachers jointly involved in detecting gaps and monitoring learner progress in targeted domains.

Box 2. A Student-friendly Rubric: Checking What I Know in Long Division (LD)

I can:

- Say in my own words what the LD problem is asking me to do
- Read a story problem and set up the LD and/or other operations needed to solve it
- Identify the “dividend” and “divisor” in an LD problem
- After the LD problem is set up, know how to start the algorithm (find digit with the highest place value in the dividend)
- Recall and use multiplication tables accurately when doing long multiplication and LD
- Recall place value concepts and give place value labels for digits in whole and decimal numbers, while doing long multiplication and LD
- Follow the steps of the LD algorithm correctly (for e.g., students use a mnemonic from class--DMSBAR)
- Repeat the LD algorithm until I get a remainder that cannot be further divided (grade 5)

OR

Continue the LD algorithm so I can express remainder as a decimal of the quotient (grade 6)

- Check the LD answer with “backwards” operations
- Explain what the answer means in my own words
- Think-aloud and write the steps of LD
- Think aloud and explain the answer and parts of the answer in a story problem
- Keep my scratch work neat (so that I don’t get lost)
- Not be afraid to make mistakes
- Not give up before I am finished
- Teach LD to a friend
- Make connections between fractions, decimals and division/LD operations

Ideally, PALD data must also be generated proximally to learners, while instruction and learning is *in process*--through probing and questioning during instruction, or appropriately designed home-work, class-work, pre-unit or mid-unit exercises. A key part of the PALD process is the timeliness and quality of follow-up actions taken by teachers, namely, *mediation* and *practice opportunities*. Box 3 provides a sample of actual actions taken by a PALD teacher during a 3-day unit on the LD domain with the class used for the case study.

Box 3. A Sample of PALD Activities during a Mini-unit on the LD Domain

1. On Day 1, the lesson began with a pre-assessment administered by the teacher. The reflections showed the following thread of reasoning in the PALD lesson plan:

The first assessment I used was an integrated, real-life math task of mid-level difficulty on the domain continuum (coconut cookie problem). It was the type of task on which we were shooting for mastery. My plan was to administer that task right at the beginning and let the instruction flow from what we found there. The task was a 2-step, real-life “word” problem involving several embedded competencies and concepts in long division, but no decimals. We were aiming for solving with decimal numbers, but it was too early. Students would have to evaluate their answers after the long division procedure was complete to see if the question in the problem was fully answered. If not, they would need a couple more operations to complete the problem-solving. Interpreting different parts of the answer in the problem context was important for obtaining a high score. The assessment was accompanied by some open-ended questions probing students’ reasoning, thinking and feelings of confidence regarding long division.

2. Error analysis and review of the responses by the teacher(s) at the end of Day 1 showed an array of errors, but one specific area with high levels of learner confusions/blocks.

[88% of the class of 17 students made] place value mistakes while doing the algorithm—placing part solutions in wrong places and bringing down the wrong number from the dividend or getting stuck because of misplacements. Many were giving up after this misstep—leaving blank, scribbles or unfinished responses.

3. The teacher’s reflection and inferences about that subset of children’s cognitive and mediation needs were as follows:

[Need to] deepen understanding of what division is and how place value is related to bring down step of division; [they] need better understanding of why we do long division of a 3-digit number in parts, starting from the digit with the highest place value; why do we need to stay in columns when we bring down? What do digits in particular columns represent?

4. Mediation techniques that the teacher documented using were:

Use labels, H(undreds), T(ens), O(nes), 1/10, 1/100, etc., as place value markers over columns of digits in the dividend (to correct place value errors).

*Think aloud the step in the algorithm and say why it was needed (to correct math reasoning errors), e.g., *If 5 fits into the 7 in 700 one (1) time, we place the 1 or that part of the answer in the hundred’s place of the quotient. Let us think--Why? Yes...Because it means we are multiplying 5 x 100 to get 500, so—now we can see what’s left over when we subtract it from 700.. See?.. the 7 in 700 is also in the hundred’s place.**

*Prompt and probe during recap/demonstration of problem-solving (to build math concept understanding and correct vocabulary issues), e.g., *When we do long division, we think of the divisor as a “grouped” number or a set. When you look at a real life problem, we can look for word clues in the story to see if there is a “grouped” number or a set. In the cookie problem, what is the grouped number that is most probably the divisor? Is it “packets” of cookies? So, what do we divide 119 by—12?**

5. Results from the follow-up practice assessment at the end of Day 1 (similar task with Orbit chewing gum packets, with 14 to a pack) showed that the error rate on that concept/skill dropped to 50% in one day. PALD approaches continued through the end of the school year. At the end, students were able to make nuanced statements about their own learning using the math vocabulary, such as:

[Problems] 14 & 17 they were the hardest because you had to divide numbers by 3 digit numbers and 2 digit numbers.

Theoretical Principles underlying the PALD Approach

A basic psychological premise on which PALD stands is the notion that cognitive and affective capacities in relation to gaining expertise in given domains of performance, can be modified through proximally-conducted, dynamic assessment and mediation. These principles are adopted from a body of compelling work by Feuerstein and his colleagues in the late 1970s who were able to show intellectual gains with this method in mentally retarded adolescents in Israel (Feuerstein et al, 1979).

The ideas on cognitive modifiability are further reinforced by more current research in the neurosciences suggesting that intellectual, affective and emotional capacities in humans are not static, but can develop and change over one's lifetime (Goleman, 1994; 2007). Interventions that forge and nurture mental connections or alter neural networks, can change behavior patterns in humans, resulting in changes in our cognitive, emotional and social response patterns and functioning.

Additionally, recent research on how individuals learn and gain expertise in particular domains, synthesized by Pellegrino, Chudowski & Glaser (2001), reinforce Feuerstein et al's (1979) conclusions and broaden the base from which learning-centered assessment in classrooms can be approached. The key elements are that (see Pellegrino, Chudowski & Glaser, 2001, p. 61-63):

- A person's prior knowledge, if made relevant to solving novel tasks/new kinds of learning, helps construction of new knowledge and strengthens cognitive bonds
- Scaffolded and facilitative instruction can help build new mental schemas, when learning stalls or connections are not made by learners;
- Problem-solving abilities and related learning are enhanced when learning tasks are meaningful and situated, rather than de-contextualized;
- Learning and problem-solving are integrative processes, where learners typically draw on multiple cognitions--skills, concepts, and higher order abilities—as relevant to a given task;
- Learners' affective states and intellectual capacities are interdependent

- Metacognitive capacities (i.e.,—the ability of learners to think introspectively and appraise their own understandings and grasp of a task/performance domain as they learn) facilitates learning and development in given areas.

The *proximal* and *dynamic* characteristic of assessment for learning, surfaced as a necessary and critical element in modifying learner outcomes in the above body of literature on cognitive and developmental psychology.

The above principles were complemented with selected ideas from the literature on domain-referenced testing, a close relative of criterion-referenced testing (CRT) methodology in educational measurement, and from error analysis research in cognitive psychology (see Millman & Greene, 1989; Nitko, 1989 for details on historical traditions of CRT; Nichols, 1994).

Documentation on how traditional CRT and domain-referenced testing has played out in current state standards-based testing programs under NCLB, is also now available (see Linn, Baker, & Betebenner, 2002). However, as already elaborated, PALD domains are different than traditional connotations of test domains in domain-referenced test design.

PALD tests are designed from ordered domains, with the domain specification process employing a backwards design procedure described. Additionally, the use of situated and integrative tasks, and incorporation of probing and scaffolding within test design, are key principles incorporated into the developmentally-diagnostic test design methods of the PALD project that are not found in traditional CRT.

Further, traditional connotations of the term “diagnosis”, as used in standardized, skill-based diagnostic inventories are also different from PALD. Typical diagnostic tests, such as the commercially-distributed Key Math Diagnostic Arithmetic Test, yield point-in-time indices of a person’s proficiency on different competency areas, showing profiles of strengths and

weaknesses of examinees by topical domains that report on status of an individual at a particular time, such as Although the information is intended to have formative utility in instructional contexts, in contrast to the dynamic PALD approach, these have been described in the literature as *static* diagnostic tests.

In terms of the PALD instructional design, hierarchical or progressive lists of expected outcomes and competencies with associated task analyses as the domain frameworks, are found in the earlier work of Robert Gagne (1970). Such tests were often embedded in computerized instructional modules, and tapped into pre-requisite skills and knowledge of students pertinent to their attaining mastery of a domain. Feedback loops from such formative assessment were meant to be closely tied to instruction, informing teachers on the effectiveness of their instructional strategies as well as the adequacy and appropriateness of goals of instruction (Gagne, 1997).

Error analysis research offered ideas for building in the dynamic diagnosis and proximal data use aspects of the PALD approach for teachers, as well as the test design elements that we have labeled as “developmentally-diagnostic assessments”. Such work has been based on qualitative and mathematical modeling of examinee error patterns has been conducted by both measurement researchers and cognitive psychologists for some time now. Predominantly, the research has relied heavily on large scale tests like the SAT or TIMSS. This fact led some researchers to underscore the need for more as “cognitively-diagnostic” assessment design ideas (see Nichols, 1994). Error analysis was never before directly connected with teachers and pedagogy.

Method

We now move to describing in detail our case and case study methodology. In this section, we provide our rationale for case and domain selection, characteristics of the teacher and students in the case (along with two selected reference groups), the design specifications for developmentally-diagnostic LD tests and empirically-determined properties of test items, and discuss the rationale and procedures we used for the mixed-method analyses of the selected classroom of students.

Case Selection

Of the 14 participating PALD classrooms in the research project, we chose one class that met three criteria. These were:

- (1) Students had been exposed to home-room teachers who were PALD-participants in both grades 5 and 6 (or for two continuing years), and who were also their primary math instructors at school;
- (2) Teacher participants (and indirectly, the class) had received a majority of the PALD intervention/coaching program and supports, including a 3-day demonstration unit on PALD practices that focused on the LD domain; and
- (3) There had been little or no attrition of students in the class over the two year period, with the number of students with intact data from both Grade 5 and 6 LD assessments.

The class of 17 students met all these criteria. Only one student had been lost due to mobility, leaving 16 cases for analysis. In this case, although the class had two different teachers in grades 5 and 6, they were both PALD-trained and continuing participants in the project.

Selection of Mathematics Domains

Although the PALD intervention can be applied to any subject area, the present NSF-funded project focused on mathematics. We incorporated three units linked to local mathematics curriculum frameworks (that were scheduled to be taught concentrically in both grade levels), into the PALD intervention program. They were units on: long division (LD), geometry (angles and triangles), prime factorization in the first year, and in the second year, algebra. When we began the research, the leaders, staff and a local consultant informed us that LD is typically taught to students at the district from grade 3 onwards. However, it is a topic with which students continue to struggle even in grade 6. This information became a main motivation for focusing on long division for in-depth testing of the PALD approach.

Student and Teacher Characteristics

Case. The *teacher* of the class in grade 6 was female and White. She held a Master's degree in elementary education and had 9 years of teaching experience in all. She had the reputation of being a good teacher, displayed an assertive, firm and confident classroom style, with an outspoken manner outside. She was also the teacher union representative at the school. The school had a good achievement history in the district on standardized tests.

The *student group* of 16 consisted of 3 (19%) Asian, 9 (56%) Black, 3 (19%) Hispanic, 1 (6%) White children, whose birth dates fell between January, 1995 through March, 1996. Of these, 7 (44%) were girls. None were coded as English Language Learners (ELL) or Disabled on the district database, but 9 (56%) were on the free/reduced lunch program (an indicator of poverty).

During visits and the demonstration lessons, we found the class to be an active and animated group. Students were not shy at all with strangers in the classroom. They soon became aware of the purposes of our visits and, except for a couple who needed constant vigilance to

stay on task, fully participated in lesson-relevant interactions. Once in a while the class became noisy, and their teacher needed to intervene to help manage classroom proceedings.

Reference Groups. To contextualize the results of the case in terms of the overall change in LD domain scores, we also looked descriptively at two groups of Grade 6 students: those who had had PALD-trained teachers for two continuing years in the four participating schools, and those who had Non-PALD teachers for two continuing years in the same schools. It should be noted here that the purpose for using the reference groups was *not* to examine comparatively whether the class in the case study performed better or worse than the other groups, but to better gauge what the typical change pattern was for same-aged children exposed to the same curriculum, both with and without the PALD intervention. The characteristics of the reference groups, described below, were very similar to the district demographic profiles for students and teachers at the elementary level in terms of ethnicity and poverty.

The group of PALD teachers, including the teacher in the case, were 5 in number. They were all White; all held Master's degrees; 2 were female; and their teaching experience ranged from 4-11 years.

The total PALD student group at the end of grade 6 consisted of 8 (12%) Asian, 42 (62%) Black, 13 (19%) Hispanic, 5 (7%) White children. The PALD students as a whole were older than the children in the case, with dates of birth extending from December, 1993 through March, 1996. Of these, 33 (49%) were girls; 7 were coded as English Language Learners (ELL) and 3 as Learning Disabled in the district database; 34 (50%) were on the free/reduced lunch program.

The Non-PALD teachers, including one substitute teacher, were 15 in number. They were also all White and had Master's degrees; 12 were female. Other than the substitute teacher with no experience, teaching experience here ranged from 4-23 years.

The Non-PALD student group consisted of 2 (1%) American Indians, 16 (7%) Asian, 143 (65%) Black, 38 (17%) Hispanic, and 22 (10%) White children. These students were also older than the children in the case, with birthdays extending from May, 1993 through November, 1995. Of these, 122 (55%) were girls; 27 were coded as English Language Learners (ELL) and 15 as Learning Disabled and 4 with other disabilities on the district database; 84 (38%) were on the free/reduced lunch program.

Data Source

The data source consists of student responses to structured-response and open-ended items on two different but parallel forms of developmentally-diagnostic tests tied to the LD domain, that were administered to the group of students at the end of grade 5 (Year 1, or with one year's exposure to a PALD-trained teacher) and end of grade 6 (Year 2, or with two years' exposure to PALD-trained teachers), respectively. Each of the 16 children in the case had a pair of tests that showed a picture of their development over time. As indicated earlier, Appendix B shows the details of the test design specifications, sample items, and estimated difficulty and discrimination levels of individual items and integrated item sets, in Years 1-2.

Test Design Properties. Because the test design component is integral to the PALD model and a necessary part of generating the type of data we think is necessary to support learning and learner development, we conducted some validity examinations with the test data using the class and the combined reference groups. The LD tests in both grades were parallel in terms of math content and skills tapped, with the Grade 6 test extending the application of LD procedure with decimal numbers. There were four components in each, labeled as Parts 1-4, that are organized progressively in terms of intended complexity. The first item set is described in detail to illustrate how the underlying test design principles were applied (see Appendix B).

Part 1 is an item set. It presents a numeric problem where students have to apply the LD procedure. It starts with structured items probing students' understanding of the LD operations, and taps into some vocabulary knowledge situated in the problem presented (using math vocabulary and language is emphasized in district's curriculum that was our research site.) The probes lead students to the problem to be solved, presented in an open-ended format. The questions are presented in a straight forward manner.

As during PALD implementation in the classroom, probing is "embedded" in the item set (Part 1). The items are "situated" rather than de-contextualized, in that a problem is presented first to which the probes directly refer (thus, math vocabulary and concept understanding are not tested separately but tied to a problem that students have to ultimately solve). Both structured and open-ended formats are used; the former is treated as a form of scaffolding, as the prompts are intended to draw the examinees' attention to specific aspects of the problem that they are expected to attack thereafter.

If Parts 1-4 in Appendix B are examined together, the cognitive/learning science principles incorporated in the test design are as follows: where possible, use of contextualized, situated questioning; use of embedded probing; use of inter-connected item sets that become progressively more complex; more difficult tasks calling for integrated use of multiple kinds of concept knowledge, operations, rules/procedures, or problem-solving strategies; and the possibility of both analytic (part-scoring) and holistic scoring on individual and related sets of items. The items are intended to guide the students towards thinking about and attacking the problem in a way consistent with the way they might typically be taught. The tests can be used in a grade-free manner, following children as they develop in a domain, irrespective of their grade level at school.

Psychometric Evidence. During construction, the tests and scoring rubrics were reviewed for content validity by teachers, school district curriculum specialists and math education consultant on the project, along with members of the research team. As the overall LD domain score was used to make certain inferences about children's change, and the expectation was that all items would homogeneously tap into the same proficiency area, we also examined validity and reliability with some traditional psychometric methods.

The total LD domain scores yielded internal consistency reliability estimates of .84 (grade 5) and .85 (grade 6), shown in Table 1. For the end-of-year scoring, raters were trained to use the rubrics with practice test papers in workshops. The inter-rater agreement on individual items/item sets in the first year ranged from .58 to .83, with a median value around .70. The rubrics and procedures were further examined on the basis of the inter-rater study and workshop discussions, and the scoring process improved considerably in Year 2 based on rater reports.

In terms of convergent validity with external measures, the LD total scores correlated with scale scores of the New York state standardized math tests in the order of .62 in grade 6. They correlated with internal PALD project measures in the order of .52 (geometry domain test scores), .55 (algebra domain test scores), and .38 (scores from a self report survey tapping math self-efficacy). These results together affirm the meaningfulness of the LD scores in terms of tapping some overlapping math capacities and math-related affect. The convergent validity estimates were obtained from a data set of 381 cases with complete data on all variables, collected in Year 2 of the PALD project.

The above was necessary evidence for us to support use of the LD parallel tests. Of more direct relevance to the PALD design, however, was the validation of the logic for ordering tasks by complexity, using empirical difficulty and discrimination statistics on individual items and

ordered item sets. Would the estimated difficulty increase as intended, with items/item sets designated as more complex? Also, on individual items/item sets, would the estimated difficulty change between grade 5-6, following an added year of student exposure to math teaching and PALD? The overall results for the class of 16 are shown in Appendix B. Item difficulty here refers to percent of students who gave a correct response to an item. In the case of items with rubrics yielding data on 2-7 point scales, we used the mid-point of the scale (rounded) to obtain item difficulty estimates. The item discrimination indices are adjusted item to total score correlations.

With a couple of exceptions (discussed in-depth in the Results section from a student learning angle), the general ordering of tasks was affirmed. However, the rather high and positive discrimination indices in the item statistics also raised a red flag, suggested that within the class there were major differences in levels of expertise and performance on several items/item sets, when high and low performers were sorted. This prompted a need for more in-depth examination of children's work.

Analysis of Student Responses: Cognitive Pathways of Learners

To begin the second and main phase of the analysis, children were sorted at three different levels of performance at the end of grade 6, based on the distribution of total LD domain score using total project sample's testing results on the LD Test in 2007 (N=698). We used the larger sample to derive reliable cut-scores. The three developmental levels (1-3) were based on percentile ranks of 33 (raw score 27) and 66 (raw score 36) on the Year 2 test.

Applying these cut-scores to the class for this study, we had the following breakdown:

Level 1= 5 students (lowest performance level)

Level 2=4 students

Level 3=7 students (highest performance level).

We then went back to 2006 (end-of-grade 5) to retrieve pairs of tests for each child, reflecting their status and change on the LD domain across two points in time.

We use three methods to analyze the class' data. First, the quantitative analysis uses descriptive statistics, including means, standard deviations and skewness measures to examine distributions of the class and the two reference groups over time. Changes on the overall domain score are documented using pooled standard deviation units. For the class as a whole, we also examine the item analysis results against the concepts/abilities tapped by particular item sets/items. Then, we take a look at the three developmental levels and study differences in patterns of change on each part of the LD domain (Parts 1-4).

Second, we support the above qualitative error analysis of the same children's work based on expert reviews from cognitive psychology and math education perspectives. The error analysis compares qualitative and quantitative summaries of errors made by children organized at three developmental levels at the end of Year 2, without the scoring rubrics.

Third, we compare and triangulate the results of all three sets of analyses to map cognitive pathways of learners at three developmental levels. Cognitive blocks and pedagogic needs of learners are inferred.

Quantitative Case Study Results

Quantitative Analysis for Whole Class. Learning Change based on LD Domain Scores

This section will describe findings displayed in Tables 1-3 and Figure 1. We begin with Tables 1-2, showing the overall performance of the class at the end of grade 5 and grade 6, on the LD domain score. At the end of grade 5, the class' distribution had a strong negative skew (-

.67) as may be expected with mastery-oriented approaches to teaching and assessment, like the PALD. This meant that a majority were doing well, but a few were behind. The class' mean score was 23.72 (SD=10.67), and possibly influenced by low scorers. As shown in Table 2, the class mean was placed .18 standard deviation units *below* the mean for the pooled sample in grade 5.

The PALD reference group's distribution had no skew, and the mean was a little higher in grade 5 at 24.12 (SD=11.47) than our case. The Non-PALD reference group in grade 5 had a smaller negative skew at -0.437 than the case, and the mean was well above both the class and the PALD group at 26.24 (SD=11.17).

In grade 6, the class' distribution turned around in sharp contrast to both the reference groups. The negative skew was higher in value at -.87, with a mean score of 32.12 (SD=14.66). The PALD reference group in grade 6, which included the case, had a mean of 26.02 (SD=10.52), in contrast. The PALD reference group had a clear negative skew of -.645 in the second year, suggesting that there were higher numbers of students at the top now, with greater measured levels of domain mastery. The Non-PALD had a slightly higher mean than the PALD group of 26.47 (9.34), with the skewness measure again at -.435. Both reference group means in grade 6 are lower than the class' mean, when the case data are isolated.

In Table 2 we see that the class' performance in pooled standard deviation units in grade 6 is +.59 SD units, relative to the pooled sample of 232 students. The class' gain on the LD domain registers at over +.40 SD units. The two reference groups do not look different from each other, both located at the pooled group mean, but the larger PALD group as a whole registers a small gain at +.11. The above analyses also showed that change statistics, rather than point-in-

time mean estimates, are better for gauging subtle and large shifts in overall domain mastery. Comparative mean analyses mask the picture on learning development.

Quantitative Analysis for Whole Class. Learning Change based on Item Analyses

Looking at the results in Appendix B more closely, it becomes possible to look at how the class performed based on concepts/cognitive processes tapped by particular items/item sets. Because the class N is 16, item statistics may not be reliable for evaluating test properties. However, they are employed here to study where learning changed for the class and where it appeared to stall. The very high positive discrimination indices at the item level, particularly, explain the reason for the sharp negative skew in the grade 6 distribution for the class/case under study.

We find in Appendix B that the first LD concept understanding probe (#1) is relatively easy at the end of grade 5 with a difficulty estimate at .75. It discriminates positively between high and low performers at +.26. As expected following another year of intervention, this type of item becomes much easier for the class as a whole in grade 6 with difficulty at .94. But, the type of item now discriminates more between high and low performers with a discrimination value of +.43.

In contrast, the open-ended item tapping the same kind of LD understanding (asking for a restatement of the problem), is considerably harder for the class in grade 6 with difficulty at .75, than a structured item tapping the same ability (#1). Although these were expected patterns, we find that loss of scaffolding in an open-ended item, makes the task harder for many children in the class, yielding a discrimination value of +.63.

Also in contrast, the item difficulty the LD vocabulary item (item type # 3, tapping recall of concept knowledge) remains unchanged over time, and is relatively easy in both years at .87-.88. As intended in the same item set, the application of the LD procedure (item type #6, Part 1,

numeric problem) in grade 5 is more difficult than the preceding items tapping concept knowledge and understanding at .63. The same item becomes easier in grade 6 at .75. In both years, we find that the application task is highly discriminating in terms of performance of students who can and cannot do LD (D values of +.83 and +.85).

As intended, the most difficult item for the class was Part 4 (real-life problem requiring use of decimals) at .50. This item was added to the developmental domain in grade 6. It also discriminated positively and strongly at +.69.

There were two unexpected results in terms of how the tasks (test domain) interacted with the class over time, shown in test results to Part 2 and Part 3 in Appendix B. On Part 2, we see the items becoming *slightly more difficult* in grade 6 (drop of difficulty from .87 to .81). On Part 3, dealing with real-life problem-solving, a major target set for grade 5, we see *no* change in the overall item difficulty or discrimination values (.75 difficulty, +.94 discrimination). This last finding suggests that while the class achieved some level of mastery in grade 5 on this type of task, there was no development beyond that point.

Quantitative Analysis by Developmental Level. Learning Change on Items/Item Sets

To identify learning pathways, Table 3 may now be examined by looking across the three developmental groups' performance by item/item set. The table identifies with bold and italics when changes occur suggesting gains from grade 5-6, on each item set (Parts 1-4) of the LD domain.

On the initial item probes in Part 1, we see some gains made by the children at the lowest level. This group (Level 1) shows difficulty with the open-ended item asking for restatement of the LD problem even in grade 6 (only 40% get it right), but they are able to handle similar probes

when scaffolded with a structured item format. The above conceptual difficulty disappears at developmental Levels 2 and 3, irrespective of item format.

On the multi-step LD problem in Part 1, we find *no change* in the distributions in grade 5-6 at Level 1. However, we do see shifts at Levels 2-3 between the two grades, with *all* the children becoming successful in grade 6 (2 children were unsuccessful at grade 5).

A similar pattern is seen in shifts on Part 2, dealing with checking of LD solutions. No change is seen in the lowest level children. The children at higher developmental levels have reached a level of mastery in grade 5; and this level is sustained in grade 6. The qualitative reviews of responses next, suggest why the lowest level children remain stagnant between the two grades, and the transitioning cognitive state relevant to LD mastery.

On Part 3, we see a small movement at Levels 2-3 to reach high levels of mastery in grade 6. On this item set, one child at Level 1 shows a movement towards mastery.

Part 4, dealing with use of decimals during solving real-life LD problems, is clearly found to discriminate between developmental Levels 2-3. The children at highest level have reached mastery on this, although this material was introduced in grade 6; however, both lower level groups are still unsuccessful.

Figure 1 displays the change in the skew in the shape of distributions by developmental level on the overall LD domain, an indicator as to whether students were moving towards mastery, and becoming more homogeneous. The distribution is *positively skewed* at Level 1 in both years (most students at the bottom). Except for a small change in the middle bar moving right, or towards mastery, there is no significant change. There are visible changes at Levels 2 and 3, where in grade 5 we see negative skews, leading to more homogeneous distributions at the higher end of the scale in grade 6.

Qualitative Review of Selected Student Responses

Figures 2 and 3, representing samples of student responses at Level 1 and versus Level 3, now provide an added picture of mental blocks of a child who is unsuccessful and one who is an expert on the LD domain in Year 2. They suggest how learning might shift from one level to the next, with appropriate mediation techniques.

We see in Figure 2 the paper of the only child who was partly successful on the real-life problems at Developmental Level 1 in grade 6. She starts the LD algorithm, but cannot continue it and reverts to drawing groups of circles to represent the divisor (clusters of 12 circles). She manages to add enough clusters to get 167 (the dividend) and finds the remainder. This approach to doing division is taught in schools in lower grades when it is first introduced. The student is unable to progress beyond the concrete modeling approach to solving LD problems, despite being taught the LD algorithmic procedure at school. We see she gets the right answers. She interprets the answers in context. But she makes no connection at all with the LD algorithmic process and her approach to division. Her interpretations are thus partially off-the-mark.

The paper in Figure 2 was chosen because it suggests a transitional stage in development between learners who are completely stuck with LD at Level 1, and learners at Level 2 who appear to have mastery in this type of problem. If we see Figure 3, in contrast, it is clear that this child is at the highest developmental level on the LD domain.

The steps and details in working out the LD procedure are clearly displayed, scratch work shown, and the verbal explanations are on point. Along with this we see the stronger ability of the student to explain steps in writing---“thinking aloud” and communicating the process. We thus become instantly aware that we are looking at the work of an “LD expert” in grade 6.

Both the quantitative and qualitative reviews of student responses were revealing, and fit together in terms of telling a story about how the students developed over the period of our study. However, we still needed a pedagogical perspective on learning development, keeping in mind how mathematics, and LD in particular, is taught to learners at the intermediate level. The next section provides these results. (*Note to Readers: Detailed analyses in this section forthcoming from our co-author, a cognitive psychologist*).

Qualitative Analysis from a Math Pedagogy Perspective

Domain Description from a Teaching Perspective

The focus of the LD domain is on the application of the long-division algorithm to the solution of real-life division problems. Although there are two categories of division problems, partition or quotient (sometimes referred to as measurement division), the algorithm is usually taught in terms of quotient. The reason for this is that in a quotient problem you are seeking the number of groups of a given sized whole. The algorithm works more easily with the commonly taught concept approach of, “How many groups of this size fit into this number?” or at a deeper level of understanding, “How many groups of three tens fit into nine tens?” In a quotient problem you are given the size of the group and solving for the number of groups. The inverse (backwards) relationship of quotient to multiplication is also clear. “How many groups of size three are there in 15?” is the inverse of the multiplication concept within, “What will be the size of the whole if there are five groups of 3?”

In contrast, in a partition problem you are given the number of groups or parts and seeking the size of each part or “If I divide 15 into five parts how large will each part be? Although the algorithm works for both forms and students quickly learn that the larger number

must be the dividend, the interpretation of the quotient and remainders for problem solving are a greater challenge.

The algorithm has been traditionally taught by rote memorization of the sequence of steps (sometimes with an acronym); divide, multiply, subtract, bring down and again or repeat. More recently, greater understanding of the process has been added in the presentation of long division as a sequence of steps in which the units of each place value from the greatest to the least are divided and remainders regrouped to the next smaller place value. In either case, the ability to identify and understand the meaning of dividend, divisor, quotient and remainder is a sometimes-overlooked but necessary basic to the ability to solve a given algorithm or construct the algorithm from a word problem.

Qualitative Identification of Errors in Grade 5 Responses

Developmental Level 1. The only question answered correctly by most students was item (1), which was structured-response that asked them to choose the meaning of an illustrated division problem. The students also seemed to be able to identify the rule for division by five, but there were several errors in the facts. Three students were able to identify the first step from structured item. Although two students at Level 1 were able to set up the algorithm, all were unable to complete the algorithm correctly—even with a single digit divisor. The errors included:

- knowledge of the facts (multiplication and division)
- subtraction process
- incorrect placement of partial products as well as the sequence of steps

Developmental Level 2. All students at Level 2 identified the divisor and dividend and division by five rule. Three out of four of the students were able to identify the meaning of the

illustrated algorithm, where to begin, and correctly performed LD with a single and double digit divisor, identified the remainder and constructed the steps taken.

Developmental Level 3. The only noticeable problem encountered by the Level 3 group was in choosing that the first step was to look at the dividend from the number on the left. Only 2 students in the group of seven at this top level chose the right answer--yet they all performed the algorithm correctly. Two of the group did not precisely describe the steps they performed for the algorithm. One omitted the item and the other merely abbreviated the numerical calculations

Qualitative Identification of Errors in Grade 6 Responses

Developmental Level 1. There was some evidence of the ability to identify the divisor and dividend. Most were also able to identify the divisibility rule and choose a first step, which was expressed as a place value. The entire group, however, could not correctly construct the sequence of steps even though the names were listed. Only one student of the group was able to complete the algorithm correctly. None could construct and complete the algorithm from the problem.

Developmental Level 2. All students were able to identify divisor and dividend, division by 2 rule (most with an example), correctly perform the single digit, no remainder algorithm and with one exception list the steps in order. Interestingly, most of the students in this group were also able to accurately construct a two-digit algorithm from the problem given and correctly complete it, but they had some difficulty in identify the quotient and in constructing (in words) the meanings of divisor, dividend, and remainder. Only one student correctly checked the quotient with the inverse multiplication. The most difficulty for the group was in setting up the partition problem and converting the remainder to a decimal.

Developmental Level 3. The question on this test about where to begin LD was phrased in terms of highest place value, but two of this top group still made the wrong choice. All set up the

algorithms correctly and completed them with single and double digit divisors and decimal conversions of remainders. An interesting observation for this group is that when they had a LD problem with remainders they automatically converted the remainders to decimals and sometimes rounded the quotient to the next whole number. This sometimes confounded the expected explanation of the problem solution.

Inferences about Pedagogic Needs

Student responses show that the ability to correctly perform the rote algorithm sometimes seems to be separated from its connection to a problem to be solved. When performed without understanding, LD is less of a challenge. This is the pattern at Level 2. The items in the assessments for each year represented a sequence of content difficulty levels from a pedagogy perspective, with ranges including: single digit to double digit divisors, quotients with and without whole number remainders to conversion of remainders to decimals, and un-constructed to open-ended constructed responses.

For the lowest developmental level, there is a need for recursive review of the division facts, meaning of division, and steps in the algorithm as well as practice in applying the connected vocabulary and using it to describe the process. They still definitely need some individualized support using manipulatives and direct connections between these, the problems, and the algorithm.

Students at Level 2 may need more practice with setting up partition vs. quotient problems as well as decimal applications. The processes within the algorithm as well as the interpretation of the quotient as a solution to the problem are critical. Application of the inverse multiplication as a check, is also needed practice.

It is obvious that the Level 3 group has mastered the algorithm process in grade 6. Recursive review of further applications to both quotient and partition real-life problems may be helpful in firming up long-term memory and preparation for applications to algebraic expressions.

Conclusions

The case in this paper showed how children attained expertise in the LD domain, using results from a particular type of diagnostic testing and test design integral to the PALD approach, coupled with methods in mathematics pedagogy prevalent at the district research site. Results identify different ways in which children at different developmental levels "think" as they improve in long division, suggesting how the PALD approach can be used to further support their learning.

The class made striking gains between grade 5 and 6 on the LD domain. However, those at the *lowest developmental level were left behind*. Domain-specific gaps became very clear through a combination of descriptive quantitative analysis and qualitative reviews of responses; further types of mediation and scaffolding that might be effective are also suggested in the results.

In terms of cognitive pathways, we find that students at the lowest Level 1 had not mastered the LD algorithm by and large over the two year period, but were able to respond when *scaffolded*, to probes on LD understanding. There appeared to be a transitional stage whereby they perform division with understanding, but using concrete pictorials/manipulatives. To make the next cognitive connections to use of the LD algorithmic process, they need further scaffolded instruction and assessment.

The mid-level students at Level 2 appear to master the algorithmic process, but they might often do it by rote (without understanding or the ability to explain). The learning and mediation needs for this level of learning are more practice with reasoning, thinking aloud and explaining. The LD experts at Level 3 gain the ability to articulate and communicate what they think, along with masterful and clean application of the procedure.

We found a good deal of agreements in the quantitative analysis with data from the developmentally-diagnostic tests and the qualitative observations and reviews of student growth. This triangulation may be viewed as a validation of the evidence generated from the tests.

The types of quantitative and qualitative analyses illustrated are not overly complex, and may be meaningful to most teachers. Thus, a next step might be in incorporating materials from case studies such as this one in teachers' professional development materials. Anecdotally, we can report that the teacher in our case study communicated serious reservations about the time and work the PALD approach entailed in the first year of the intervention research. At the end of the second year, she specifically asked to continue using the PALD approach in the following academic year, despite the fact that the grant-funded project had reached closure and there would be no funding/supports available.

This study is significant in that it adds to the theoretical knowledge base on cognitive and assessment sciences as it impacts classroom teaching practices and learner development, successfully marrying these areas, as recommended in recent national academy reports (Pellegrino, Chudowski & Glaser, 2001). The ideas for developmentally-diagnostic test design as presented here, marry the highlighted ideas from psychological literature, with selected principles from the assessment sciences. Further research and development on the PALD model is continuing.

References

- Bloom, B. S. (1982). *Human characteristics and school learning*. New York, NY: McGraw-Hill Book Co.
- Chatterji M. & Gordon, E.W. (2004). *Improving mathematics achievement in grade 5-6 students with systemic use of proximal assessment data*. New York, NY. Teachers College Columbia University. Proposal supported by the National Science Foundation, EREC solicitation. December, 2004.
- Feuerstein, R. et al (1979). *The dynamic assessment of retarded performers: The learning assessment potential device, theory, instruments and techniques*. Baltimore, MD: University Park Press.
- Fuchs, L.S., & Fuchs, D. (2003). Can diagnostic reading assessments enhance general educators' instructional differentiation and student learning? In B. Foorman (Ed.) *Preventing and remediating reading difficulties: Bringing science to scale* (pp. 325-351). Baltimore, MD: York Press.
- Gagne, R. (1970). *The conditions of learning* (2nd. Ed.). New York, NY: Holt, Rienhart & Winston.
- Gagne, R. (1997). Mastery learning and instructional design. *Performance Improvement Quarterly*, 10(1), 8-19.
- Goleman, D. (1994). *Emotional intelligence: Why it can matter more than IQ*. New York, NY: Bantam.
- Goleman, D. (2007). *Social intelligence: The new science of human relationships*. New York, NY: Bantam-Dell.

- Guskey, T. R. (2003). How classroom assessments improve learning. *Educational Leadership*, 60(5), 6-11.
- Linn, R.L., Baker, E. L., & Betebenner, D. W. (2002). Accountability systems: Implications of requirements of NCLB Act of 2001. *Educational Researcher*, 31 (6), 3-16.
- Millman, J., & Greene, J. (1989). The specification and development of tests of achievement and ability. In *Educational Measurement* by R. L. Linn (Ed.), pp. 335-366: American Council on Education.
- Nichols, P. D. (1994). A framework for developing cognitively-diagnostic assessments. *Review of Educational Research*, 64 (4), 575-603.
- Nitko, A. J. (1989). Designing tests that are integrated with instruction. In R.L. Linn (Ed.), *Educational Measurement* (3rd ed., pp. 447-474). New York: American Council on Education and MacMillan Publishing Co.
- No Child Left Behind Act of 2001 (NCLB), *Public Law No. 107-110*, 115 Statute 1425 (2002).
- Pellegrino, J.W., Chudowski, N., & Glaser, R. (Eds). (2001). *Knowing what students know: The science and design of educational assessment*. Washington, D.C: National Academies Press.
- Shepard, L.S. (2000). The role of assessment in a learning culture. *Educational Researcher*, 29(7), 4-14.
- Stiggins, R. J. (2002). Assessment crisis: The absence of assessment for learning. *Phi Delta Kappan*, 83(10), 758-765.
- Wiggins, G. (1998). *Educative assessment*. San Francisco: Jossey-Bass.

Table 1. Performance of Class Selected for Case Study versus Student Cohorts* on Long Division (LD) Domain: Descriptive Statistics in Year 1 and Year 2 on LD Total Scores

| <u>Year/Grade</u> | <u>Group</u> | <u>Mean</u> | <u>SD</u> | <u>Max Score</u> | <u>Min Score</u> | <u>Skewness</u> | <u>Kurtosis</u> | <u>N</u> | <u>Standardized Alpha Reliability for LD Score</u> | <u>Item-level Inter-rater Agreement: Range (median)</u> |
|-----------------------------|-----------------|--------------|--------------|------------------|------------------|-----------------|-----------------|----------|--|---|
| 2006/ End of Grade 5 | Class | <i>23.72</i> | <i>10.67</i> | 34 | 3 | <i>-0.672</i> | -1.278 | 16 | .84 | .58-.75 (.70) |
| | PALD Cohort | 24.12 | 11.47 | 43 | 0 | <i>-0.058</i> | -0.981 | 55 | | |
| | Non-PALD Cohort | 26.24 | 11.17 | 43 | 0 | <i>-0.437</i> | -0.818 | 161 | | |
| 2007/ End of Grade 6 | Class | <i>32.12</i> | <i>14.66</i> | 47 | 6 | <i>-0.866</i> | -0.721 | 16 | .85 | Pending |
| | PALD Cohort | 26.02 | 10.52 | 40 | 3 | <i>-0.645</i> | -0.378 | 55 | | |
| | Non-PALD Cohort | 26.47 | 9.34 | 40 | 2 | <i>-0.435</i> | -0.553 | 161 | | |

* *Note:* PALD student cohort (includes class for case study) was taught mathematics for two continuing years by PALD-participant teachers. Non-PALD cohort was taught mathematics for two continuing years by non-participant teachers. Raw scores not comparable in Year 1 and 2 due to different test lengths.

Table 2. Grade 5-6 Gains in Pooled Standard Deviation Units on Long Division (LD) Total Scores: Class Selected for Case Study versus Student Cohorts*

| <u>Year/Grade</u> | <u>Group</u> | <u>Mean</u> | <u>SD</u> | <u>Mean z-score</u> | <u>Mean z-score Gain-Year 2</u> | <u>N</u> |
|-----------------------------|-----------------|--------------|--------------|---------------------|---------------------------------|----------|
| 2006/ End of Grade 5 | Class | | | -.18 | | 16 |
| | PALD Cohort | | | -.15 | | 55 |
| | Non-PALD Cohort | | | +.04 | | 161 |
| 2007/ End of Grade 6 | <i>Pooled</i> | <i>25.79</i> | <i>11.12</i> | | | 232 |
| | Class | | | +.59 | <i>+.41</i> | 16 |
| | PALD Cohort | | | -.04 | <i>+.11</i> | 55 |
| | Non-PALD Cohort | | | +.00 | <i>+.00</i> | 161 |
| | <i>Pooled</i> | <i>26.44</i> | <i>9.67</i> | | | 232 |

*Note: PALD student cohort (includes class for case study) was taught mathematics for two continuing years by PALD-participant teachers. Non-PALD cohort was taught mathematics for two continuing years by non-participant teachers.

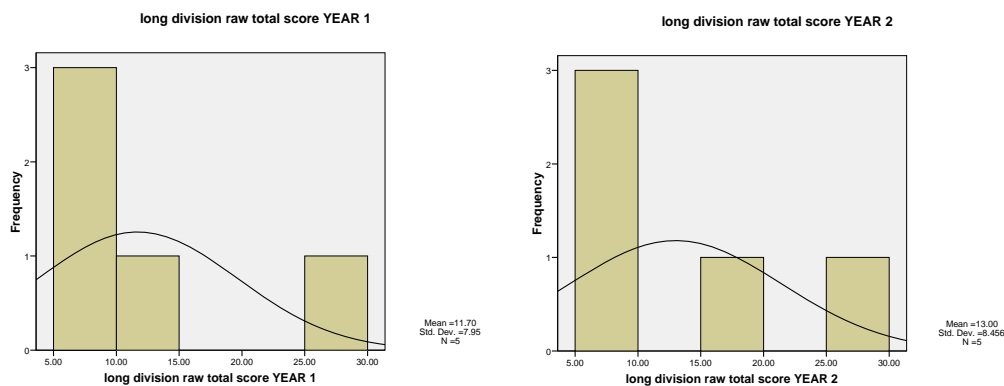
Table 3. Item-Level Performance of Sub-Groups by Developmental Level on Long Division (LD) Domain: Changes* from Year 1- Year 2

| Task Set | Task Descriptor | Score Levels | Developmental Level 1 | | Developmental Level 2 | | Developmental Level 3 | |
|----------|--|--------------|-------------------------|--------------------------------|-------------------------|-------------------|--------------------------------|-----------------|
| | | | Grade 5 | Grade 6 | Grade 5 | Grade 6 | Grade 5 | Grade 6 |
| Part 1 | Situating understanding of LD numeric problem-rephrase-MCQ | 0 | <i>f</i> (%) 2 (40%) | <i>f</i> (%) 1 (20%) | <i>f</i> (%) 1 (20%) | <i>f</i> (%) - | <i>f</i> (%) 1 (14%) | |
| | | 1 | 3 (60%) | 4 (80%) | 3 (75%) | 4 (100%) | 6 (86%) | 7 (100%) |
| | Open-ended item (0/1) | 1 | - | 2 (40%) | | 4 (100%) | | 7 (100%) |
| | | 0 | 2 (40%) | 2 (40%) | 4 (100%) | 1 (25%) | | |
| | Situating use of LD Vocabulary (1) | 1 | 3 (60%) | 3 (60%) | | 3 (75%) | 7 (100%) | 7 (100%) |
| | | 0 | 2 (40%) | 1 (20%) | 4 (100%) | 1 (25%) | | |
| | Situating use of LD Vocabulary (2) | 1 | 3 (60%) | 4 (80%) | | 3 (75%) | 7 (100%) | 7 (100%) |
| | | 0 | 2 (40%) | 1 (20%) | 4 (100%) | 4 (100%) | 1 (14%) | |
| | Situating, scaffolded understanding of LD numeric problem-start step | 1 | 3 (60%) | 4 (80%) | | | 6 (86%) | 7 (100%) |
| | | 0 | 3 (60%) | 3 (60%) | 1 (25%) | - | | |
| | | 1 | - | - | - | - | - | - |
| | | 2 | 1 (20%) | 1 (20%) | - | - | - | - |
| | Scaffolding solving of multi-step LD numeric problem; interpreting answers, using language | 3 | - | - | - | - | 1 (14%) | |
| 4-6 | | 1 (20%) | 1 (20%) | 3 (75%) | 4 (100%) | 6 (86%) | 7 (100%) | |
| 0 | | 1 (20%) | 1 (20%) | | | | | |
| 1 | | - | - | - | - | - | - | |
| Part 3 | Scaffolding-checking of LD solution; interpreting answers, using language | 2 | - | 2 (40%) | - | - | | |
| | | 3 | 2 (40%) | - | 1 (25%) | - | | |
| | | 4-7 | 2 (40%) | 2 (40%) | 3 (75%) | 4 (100%) | 7 (100%) | 7 (100%) |
| | | 0 | 3 (60%) | 1 (20%) | - | | | |
| Part 4 | Unscaffolding solving of multi-step real-life problem needing LD and other operations with 2 and 3 digit divisors; interpreting answers, using language | 1 | 1 ((20%) | 2 (40%) | - | | | |
| | | 2 | - | | - | | | |
| | | 3 | - | | - | | | |
| | | 4 | 1 (20%) | | 1 (25%) | | 1 (14%) | |
| | | 5-9 | - | 2 (40%) | 3 (75%) | | 1 (14%) | |
| | | 10+ | | | | 4 (100%) | 5 (71%) | 7 (100%) |
| Part 5 | Unscaffolding- solving of multi-step real-life problem with 2-3 digit divisors and decimal numbers, needing LD and other operations; interpreting answers, using LD language | 0 | Not tested | 5 (100%) | Not tested | | Not tested | |
| | | 1-4 | - | - | - | 3 (75%) | | |
| | | 5-9 | - | - | - | 1 (25%) | - | 7 (100%) |

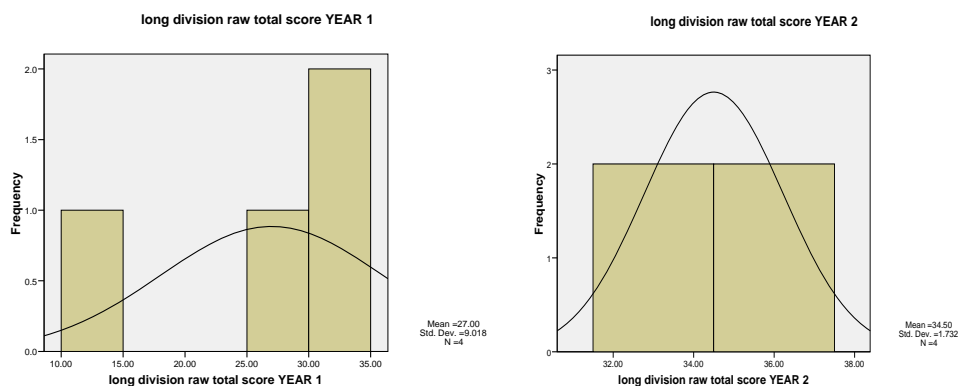
*Note: See Appendix B for examples of ordered items, design specifications, and statistics on item difficulty and discrimination in Years 1 and 2.

Figure 1. Case Study of Class Exposed to PALD-Trained Teachers for Two Years: Changes in Mastery of Long Division at Three Developmental Levels

Developmental Level 1: Change from Year 1-Year 2 on Long Division Domain (N=5)



Developmental Level 2: Change from Year 1-Year 2 on Long Division Domain (N=4)



Developmental Level 3: Change from Year 1-Year 2 on Long Division Domain (N=7)

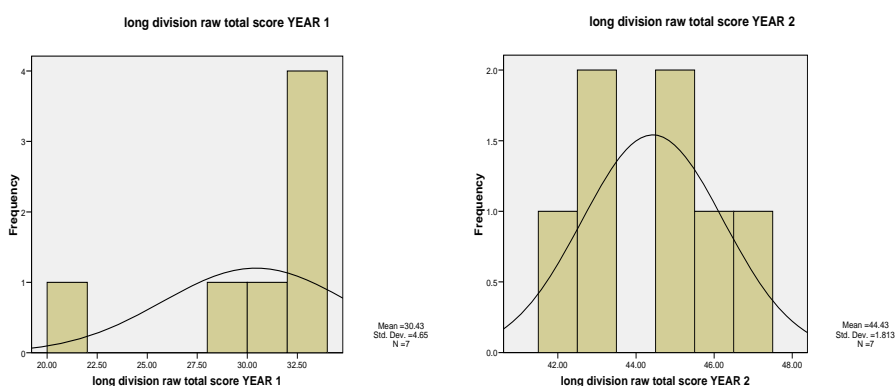


Figure 2. Cognitive Processes of a Learner at Developmental Level 1: Pathways to Mastering Long Division

Teachers College, Columbia University – PALD 2007
Developmental Assessment: LONG DIVISION

For Questions 14 through 16, read the following story.

There are 167 seventh graders who will attend the annual school picnic. Mrs. Tabella, who is in charge of the picnic, wants all the students to sit at picnic tables. Each table can seat 12 people. How many tables does Mrs. Tabella need to seat all the seventh graders? Set up the problem clearly and show all your work to get the answer.

14.a.
$$\begin{array}{r} 13 \overline{)167} \\ \underline{12} \\ 47 \\ \underline{36} \\ 11 \end{array}$$

13 tables
11 people left



Teachers College, Columbia University – PALD 2007
Developmental Assessment: LONG DIVISION

14.b. What is the quotient in the problem (#14)? 13

14.c. What does it stand for in the story? Say in words.
how many tables there are.

14.d. What is the divisor in the problem (#14)? 12

14.e. What does it stand for in the story? Say in words.
the number I am going to divide by

14.f. What is the dividend in the problem? 167

14.g. What does it stand for in the story? Say in words.
It stands for how many time I am going to divide, and how many times it is going in to.

14.h. Is there a remainder? yes

14.i. What does the remainder stand for in the story? Say in words.

It stands for how many people are left in the story to be seated at a table.

15.a. Mrs. Tabella will need _____ picnic tables to seat 167 seventh graders.

15. b. Explain how you got the number of tables Mrs. Tabella needed. Write down the main steps.

Figure 3. Cognitive Processes of a Learner at Developmental Level 3: Pathways to Mastering Long Division

Teachers College, Columbia University – PALD 2007
Developmental Assessment: LONG DIVISION

For Questions 14 through 16, read the following story.

There are 167 seventh graders who will attend the annual school picnic. Mrs. Tabella, who is in charge of the picnic, wants all the students to sit at picnic tables. Each table can seat 12 people. How many tables does Mrs. Tabella need to seat all the seventh graders? Set up the problem clearly and show all your work to get the answer.

14.a.

$$\begin{array}{r}
 \times 13.9166 \\
 12 \overline{) 167.00000} \\
 \underline{72} \\
 047 \\
 \underline{36} \\
 110 \\
 \underline{108} \\
 20 \\
 \underline{12} \\
 80 \\
 \underline{72} \\
 080 \\
 72
 \end{array}$$

Scratch Work Here

$$\begin{array}{l}
 12 \times 1 = 12 \\
 12 \times 2 = 24 \\
 12 \times 3 = 36 \\
 12 \times 4 = 48 \\
 12 \times 5 = 60 \\
 12 \times 6 = 72 \\
 \\
 12 \times 7 = 84 \\
 12 \times 8 = 96 \\
 12 \times 9 = 108 \\
 12 \times 10 = 120 \\
 12 \times 11 = 132 \\
 12 \times 12 = 144
 \end{array}$$

Teachers College, Columbia University – PALD 2007
Developmental Assessment: LONG DIVISION

14.b. What is the quotient in the problem (#14)? 13.9166

14.c. What does it stand for in the story? Say in words.

How many tables do you need

14.d. What is the divisor in the problem (#14)? 12

14.e. What does it stand for in the story? Say in words.

How many people can fit at a table

14.f. What is the dividend in the problem? 167

14.g. What does it stand for in the story? Say in words.

How many students are there

14.h. Is there a remainder? yes

14.i. What does the remainder stand for in the story? Say in words.

How many kids don't have a table to sit at.

15.a. Mrs. Tabella will need 14 picnic tables to seat 167 seventh graders.

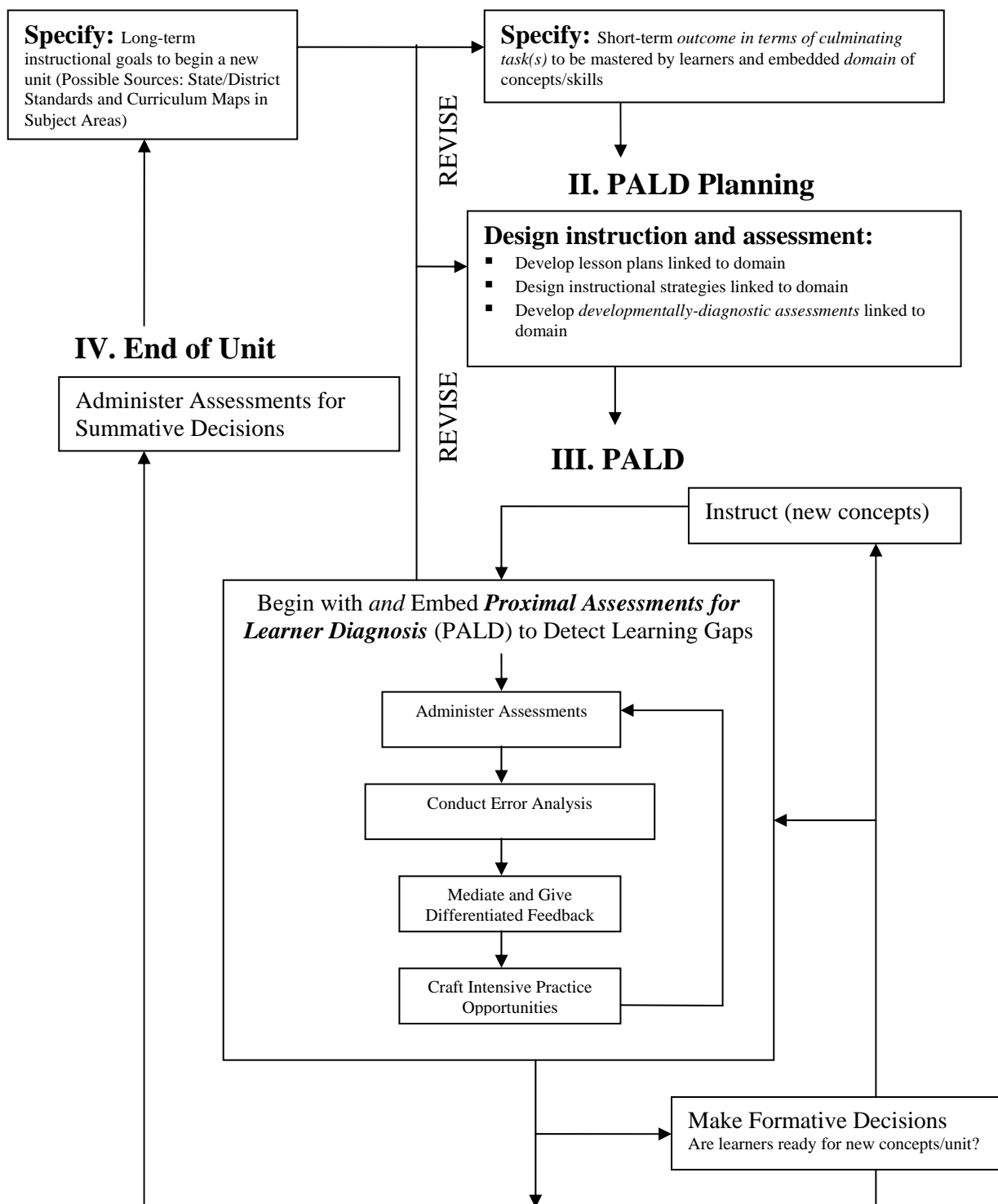
15. b. Explain how you got the number of tables Mrs. Tabella needed. Write down the main steps.

How many times does 12 go into 1, how many times does 12 go into 16, 1 then multiply, subtract, and bring down. Then do that again. When done ~~check~~ if remainder is more than 1, round up.

Appendix A

Graphic Representation of the *Proximal Assessment for Learner Diagnosis* (PALD) Model: How Teachers Work in the Classroom

I. Goal-Setting



Appendix B

Developmentally-Ordered Diagnostic Assessment in Long Division (LD): Shifts in Item Difficulty and Discrimination Statistics

| Long Division (LD) Domain: Task Examples | Task Specifications, Order, and Embedded Skills and/or Concepts from LD Domain | <u>Year 1:</u> Difficulty index (Item Discrimination in parenthesis) | <u>Year 2:</u> Difficulty index (Item Discrimination in parenthesis) |
|--|--|---|---|
| <p><i>Part 1.</i> Look at the problem below to answer Questions 1-6.</p> $5 \overline{)175}$ <p>1. The problem is telling me to:</p> <ol style="list-style-type: none"> Multiply 5 times 175 Find out how many groups of 5 are in 175 Divide 5 into 175 parts Add 5 and 175 | <p>Question or probe situated in a numeric LD problem that is presented first with 1-3 digit divisors. Problem is not solved. (Learners) interpret/show understanding of LD operation by restating what the problem is asking. Structured-response item format (scaffolded).</p> <p>Intended difficulty: Low</p> | .75 (+0.26) | .94 (+.43) |
| $5 \overline{)175}$ <p>2. Say in another way what the problem is asking. _____</p> | <p>Question or probe situated in a numeric LD problem with 1-3 digit divisors that is presented first. Problem is not solved. (Learners) interpret/show understanding of LD operation by restating what the problem is asking. Open-ended item format (un-scaffolded)</p> <p>Intended difficulty: Mid-Level</p> | <i>Not tested</i> | .75 (+.63) |
| <p>3. The 5 in the division problem is called the</p> <ol style="list-style-type: none"> divisor dividend quotient multiple | <p>Question or probe situated in a numeric LD problem with 1-3 digit divisors that is presented first. Problem is not solved. (Learners) show LD vocabulary knowledge in problem-solving context. Structured-response item (scaffolded) item.</p> <p>Intended difficulty: Low</p> | .88 (+.61) | .87 (+.15) |
| <p>5. To start to solve the problem shown, I would ask which questions?</p> <ol style="list-style-type: none"> How many times does 5 fit into 1? And if not, how many times does 5 fit into 17? What is 5 times 1? And if not, what is 5 times 17? What is 5 divided by 17? And if not, what is 5 divided by 175? Other (Explain): _____ | <p>Learners show understanding of how to start LD operation, with probe situated in a given LD numeric problem that is presented first. Problem is not solved. Structured-response (scaffolded) item</p> <p>Intended difficulty: Low</p> | .81 (+.35) | .94 (+.46) |
| <p>8. Now do the problem, showing all your work in the spaces marked below.</p> | <p>Following preceding probes into concept/operation understanding, learners apply the multi-step procedure taught, to solve a numeric LD problem with 1-3 digit divisors and dividend larger than divisor. Interpret the complete answer or parts of answer. Open-ended, constructed response item, with grouped scoring of both parts together OR with analytic rubric</p> <p>Intended difficulty: Mid- to High Level</p> | .63 (+.83) | .75 (.85) |

| Long Division (LD) Domain: Task Example | Task Specifications, Order, and Embedded Skills and/or Concepts | Year 1 Item Statistics: Difficulty index (Item Discrimination in parenthesis) | Year 2 Item Statistics: Difficulty index (Item Discrimination in parenthesis) |
|--|---|--|--|
| <p><i>Part 2.</i> For Questions 11 through 13, look at this problem:</p> $\begin{array}{r} 4 \\ 5 \overline{)23} \\ \underline{20} \\ 3R \end{array}$ <p>11. Identify the following:</p> <p>a. Quotient _____ b. Divisor _____ c. Dividend _____ d. Remainder _____</p> | <p>Question or probe situated in a numeric LD problem that is presented first, with 1-3 digit divisors and dividend larger than divisor. Problem is <i>solved</i>. Learners apply LD vocabulary, concept and procedural knowledge to check solution. Structured-response item (scaffolded).</p> <p>Intended difficulty: Mid-to High Level</p> | .88 (+.85) | .81 (+.91) |
| <p>12. Which “backwards” operations would you use to check the answer? Choose one.</p> <p>a. $5 \times 4 =$ b. $5 \times 4 + 3 =$ c. $5 \times 23 =$ d. $23 \div 5 =$ e. Other _____</p> | | | |
| <p>13. Now use your chosen formula to check the answers given. Is the solution correct? Show your work.</p> | | | |
| <p><i>Part 3.</i> Coconut cookies come in packets of 12. I would like to give 1 cookie to each 5th grader at Hillcrest Elementary. There are 119 5th graders here. How many packets of cookies should I buy so that every 5th grader gets at least one cookie? Will there be any cookies left over? Set up all the operations you need to solve the problem, solve it and explain all the parts of your answer. Show all your work in the spaces provided.</p> <p>The <i>divisor</i> in the problem is: _____ What does it stand for in the story problem? _____.</p> | <p>(Learners) apply multi-step LD procedure to solve a real life problem with 1-3 digit divisors and dividend larger than divisor. In context, they use LD vocabulary, concept knowledge, doing all other operations needed to reach solution, and interpret answer sensibly for problem, completing all steps. Open ended-item, with 5-7 point rubric. Interpretive items have structured response format (scaffolded).</p> <p>Intended difficulty: Mid- to High Level</p> | .75 (+.94) | .75 (+.94) |
| <p><i>Part 4.</i> Mr. Generosa would like to buy snacks for his students on an upcoming fieldtrip. His budget is \$25 and there are 20 students in his class. If Mr. Generosa wants to divide the money equally among his students, then how much would the snack for one student cost? Set up the problem clearly and show all your work to get the answer <u>in decimal form</u>.</p> | <p>(Learners) apply multi-step LD procedure to solve a real life problem (1-3 digit divisors and dividend larger than divisor) with use of decimals in quotient, leaving no remainder. They use vocabulary, concept knowledge, doing all operations needed to reach solution, interpret answer sensibly for problem, and complete all steps. Open ended-item, with 5-7 point rubric. Interpretive items have structured response format (scaffolded or unscaffolded).</p> <p>Intended difficulty: High</p> | Not tested | .50 (+.69) |