Developing the *Literacy Instruction Knowledge Scales* (LIKS): A Comprehensive Assessment of Primary Grade Teachers' Knowledge of Reading and Writing Instruction

By

D. Ray Reutzel

Utah State University

Janice A. Dole

University of Utah

Richard Sudweeks

Brigham Young University

Parker C. Fawson

Sylvia Read

John A. Smith

Utah State University

Becky Donaldson

Cindy D. Jones

Utah State University

Kerry Herman

University of Utah

Paper presented at the Annual Conference of the American Educational Research Association,

Chicago, April 2007

Developing the Literacy Instruction Knowledge Scales (LIKS): A Comprehensive Assessment of Primary Grade Teachers' Knowledge of Reading and Writing Instruction

Teachers influence student academic growth more than any other single factor, including families, neighborhoods, and the schools students attend (Greenwald, Hedges & Laine, 1996; Mosenthal, Lipson, Torncello, Russ & Mekkelsen, 2004; Sanders & Horn, 1994). To be effective, teachers need to have strong content knowledge and pedagogical skills (Gitomer, Latham & Ziomek, 1999; Miller, McKenna & McKenna, 1998; Moats, 2000; NRP, 2000; Shulman & Grossman, 1988). Strong content knowledge and pedagogical skills lead to "high quality teachers," which is now mandated by No Child Left Behind (NCLB) in every classroom. But, how do we prepare teachers to be "high quality teachers," and how do we know when they are? These questions have come under increased scrutiny in this early part of the 21st century (Cochran-Smith & Zeichner, 2005; Darling-Hammond, 2006; Darling-Hammond, Bransford & LePage, 2007; Goldhaber & Brewer, 1998; Levine, 2006), particularly in the area of reading (Walsh, Glaser & Wilcox, 2006). With increased scrutiny and accountability comes the need for better measures of teacher subject matter knowledge and instructional effectiveness.

Existing measures of teacher knowledge have generally taken the form of paper and pencil tests, teacher logs, and classroom observations. However, the National Reading Panel (NRP, 2000) reported that some existing assessments of teacher subject matter knowledge and instructional effectiveness lack reliability. More important, teacher knowledge measures that currently exist have yet to establish the critical linkage between teacher knowledge and student achievement. Ultimately, teacher knowledge must be linked to student achievement in order to establish the important construct validity necessary to make teacher knowledge tests useful tools

for educators. We need to know that teachers who have more knowledge produce students who achieve more. Otherwise, teacher knowledge measures are of little consequence.

The purpose of this paper is to report our ongoing efforts to develop a reliable and valid measure of teacher knowledge of grades 1-3 reading and writing instruction. In order to construct such a measure, we embarked on a four-year program of research that entailed the development and validation of a teacher knowledge measure and a classroom observation instrument that together would reliably predict reading and writing achievement for grades 1-3 students. In this paper we report our ongoing effort to develop a reliable and valid teacher knowledge measure.

Theoretical Framework

Although teacher knowledge is a difficult construct to define, theory and research have provided some indicators of the knowledge necessary to effectively teach reading and writing in the primary grades. Shulman's (1986, 1987) early distinctions between *content* and *pedagogical knowledge* and finally *pedagogical content knowledge* form the theoretical foundation for the study.

Content Knowledge

According to Shulman (1986, 1987), content knowledge is the knowledge teachers need about the content of a subject area in order to teach it. For example, teachers must know, understand and be able to manipulate fractions in order to be able to teach them to fifth-graders. Some researchers refer to this knowledge as *declarative knowledge* about a particular subject (Paris, Wixson & Lipson, 1983). Of course, the underlying assumption is that more teacher knowledge means more student knowledge. In the past, teacher content knowledge has been assessed through proxy measures such as level of education, number of content courses taken, or by paper-pencil content tests.

Assessing teacher content knowledge of reading and writing is more difficult than assessing teacher knowledge of math or science due to the complexities and the subtleties of reading and writing. Reading and writing do not have a clear body of core knowledge as do the domains of math and science. Further, readers and writers are often unaware of the content knowledge that makes them proficient. Phelps & Shilling (2004) point out, "Indeed, it is not even clear what might count as 'content' in reading" (p. 33). However, Kolis & Dunlap (2004) emphasize the importance of content knowledge because it will ultimately be the basis by which a teacher determines what is worth teaching for mastery, what is important to know and do, and what only worth mentioning.

Knowledge of content about a subject area is often addressed in teacher content standards. Beginning in the late 1980s, standards became the main criteria for determining teacher quality. The National Board for Professional Teaching Standards (NBPTS), the Interstate New Teacher Assessment and Support consortium (INTASC), and the National Council for the Accreditation of Teacher Education (NCATE) standards all provide teacher content knowledge in literacy. These documents were built through consensus and are linked to each other. All three sets of standards require that teachers have in-depth subject matter knowledge, that teachers know and use effective instructional strategies, and that teachers monitor student learning and adjust instruction accordingly (Mitchell, Robinson, Plake, & Knowles, 2001).

In its accreditation evaluations of teacher education programs, NCATE uses 33 content area professional teaching organizations as their main source of standards. The professional organization for reading is the International Reading Association (IRA). The IRA's Standards for Reading Professionals outlines the literacy knowledge and skills that teachers should have when they finish a teacher education program. The five standards are:

1. Candidates have knowledge of the foundations of reading and writing processes and instruction.
2. Candidates use a wide range of instructional practices, approaches, methods, and curriculum materials to support reading and writing instruction.
3. Candidates use a variety of assessment tools and practices to plan and evaluate effective reading instruction.
4. Candidates create a literate environment that fosters reading and writing by integrating foundational knowledge, use of instructional practices, approaches and methods, curriculum materials, and the appropriate use of assessments.
5. Candidates view professional development as a career-long effort and responsibility.

The IRA Standards for Reading Professionals were based upon professional expertise and reading research found in the three volumes of the *Handbook of Reading Research* (Pearson, Barr, Kamil & Mosenthal, 1984; Barr, Kamil, Mosenthal & Pearson, 1991; Kamil, Mosenthal, Pearson & Barr, 2000), *Preventing Reading Difficulties in Young Children* (Snow, Burns & Griffin, 1998), the *Report of the National Reading Panel* (National Institute of Child Heath and Human Development, 2000), *Theoretical Models and Processes of Reading* (Ruddell, Ruddell & Singer, 1994) and *What Research Has to Say about Reading Instruction* (Farstrup & Samuels, 2002).

Research of teacher content knowledge in the area of reading continues to grow. Emergence of the National Reading Panel Report (2000), that concluded that issues of teacher knowledge needed further research, and the Reading First legislation, which identified five principal components of instruction (Armbruster, Lehr, & Osborn, 2001), have served as catalysts for growing research of teacher knowledge of literacy development. Some research in

reading has shown that increased teacher content knowledge leads to changes in teacher behavior and classroom practices that improve student learning. For example, Moats & Foorman (2003) reported that teacher knowledge of and ability to apply concepts of phonology and orthography account for significant variance in reading and spelling achievement for primary grade students. McCutchen et al. (2002) found a direct relationship between primary-grade teachers' reading-related content knowledge and instructional practices.

Most literacy researchers would agree that, although not thoroughly defined, the content knowledge teachers need to teach reading and writing is critical to producing students who achieve. However, because many factors beyond content knowledge contribute to teacher quality (Wilson et al., 1987), Shulman (1986, 1987) has also identified pedagogical knowledge.

Pedagogical Knowledge

A second kind of knowledge that Shulman (1986, 1987) refers to is teacher *pedagogical knowledge.* Husu (2003) refers to pedagogical knowledge as "active processes by which teachers perform their duties in situations involving intense social interactions" (p. 3). This knowledge base is utilized in classroom management, routines and procedures and in general instructional practices. Pedagogy is the "art and science of teaching because it involves designing learning activities for individuals to that they learn important content themes and patterns at a deep level in a way that works for them" (Kolis & Dunlap, 2004, p. 101). Much of the research in this area has been concerned with identifying "those general forms of teaching behavior that correlate with student performance on standardized tests" (Shulman, 1987, p. 6).

There is an extensive research base for pedagogical knowledge including the effective schools research that sought to identify critical attributes of successful teachers to promote replication by others. Pedagogical research has focused on such classroom practices as teacher

expectations and praise (Brophy, 1981, 1983), questioning & feedback (Good & Grouws, 1979; Rosenshine, 1983), extrinsic and intrinsic rewards (Cameron & Pierce, 1994), classroom climate (Soar & Soar, 1976, 1979), classroom procedures (Fink & Siedentop, 1989), instructional and transition time (Anderson et al., 1979; Arlin, 1979; Durkin, 1979), and grouping (Auten, 1985; Baumann et al., 2000).

Pedagogical Content Knowledge

A third, and the most critical, component of Shulman's (1986, 1987) conceptualization of teacher knowledge is called pedagogical content knowledge. Pedagogical content or procedural knowledge "represents the blending of content and pedagogy into an understanding of how particular topics, problems, or issues are organized, represented, and adapted to the diverse interests and abilities of learners and presented for instruction" (p. 8). This knowledge is a subset of the content knowledge base; it is the particular *knowledge about teaching a subject* that teachers need to know. Pedagogical content knowledge refers to how to teach a subject to students, such as how to introduce fractions to students, which parts to include first and which parts to include later on. In reading, pedagogical content knowledge might refer to the knowledge about which letters to introduce to students early on because they are easier to learn and remember and which to introduce at a later date because they may be confusing or more complicated to learn.

Teachers clearly use both content and pedagogical content knowledge when they teach. Knowledge of literacy content provides a basic foundation of important ideas and skills for teachers. Knowledge of pedagogical practices and educational settings contribute to teacher decision making as well. Ultimately, however, teachers must merge their understanding of these components into *enacted practice*. This enacted practice of pedagogical content knowledge also

encompasses what Fenstermacher (1994) termed “practical knowledge,” the knowledge teachers generate as a result of their experiences as teachers and their reflections on these experiences. Because practical knowledge is based on experience, it is considered to be unique to the person, the context, and the content (Meijer, 2001). Although practical knowledge guides teachers’ practice, it is mainly tacit and is the least codified of all teacher knowledge (Shulman, 1987).

This study, then, attempts to capture four types of knowledge conceptualized by researchers—content knowledge, pedagogical knowledge, pedagogical content knowledge and enacted practice. The enacted practice is seen in teachers’ day-to-day handling of reading and writing instruction in the classroom. Thus, it is not only what teachers know about the content and about pedagogy, but also what they do in the classroom with students as they try to impart that knowledge to students.

The Current Study

Using Shulman’s conceptual framework (1987), this study examined a comprehensive view of teacher knowledge, one which not only values teacher content knowledge but also embraces that the pedagogical content knowledge teachers possess which allows them to understand and convey content or domain knowledge effectively to others. In addition to Shulman’s (1987) conceptual framework for describing teacher pedagogical knowledge, this study was guided specifically and conceptually by three lines of recent research related to teachers’ knowledge of primary-grade reading and writing instruction. The first of these three lines of research focuses on exemplary teaching of reading and writing in the primary-grades (Baumann, Hoffman, Duffy-Hester & Ro, 2000; Hoffman & Pearson, 2000; Pressley, Allington, Wharton-McDonald, Collins-Block & Morrow; 2001; Pressley, Wharton-McDonald, Allington, Block, Morrow, Tracey, Baker, Brooks, Cronin, Nelson & Woo, 2001; Taylor, Pearson, Clark &

Walpole, 2000). The second line of research focuses on measuring reading teacher content and pedagogical knowledge (Asselin, 1997; Bos, Mather, Dickson, Podhajski & Chard, 2001; Cunningham, Perry, Stanovich & Stanovich, 2004, McCutchen, Harry, Cunningham, Cox, Sidman & Covill, 2002; Moats & Foorman, 2003; and Phelps & Schilling, 2004.) The third line of research centered on evidenced-based practices in reading and writing instruction (Flower, 1994; Flower & Hayes, 1980; Hayes & Flower, 1980; McCardle & Chhabra, 2004; National Reading Panel, 2000; Snow, Burns & Griffin, 1998).

The study followed the recommendations of previous researchers to use a broad-based approach to measure teacher knowledge of literacy (Cunningham, Perry, Stanovich & Stanovich, 2004; McCutchen, Harry, Cunningham, Cox, Sidman & Covill, 2002; Moats & Foorman, 2003; Phelps & Schilling, 2004). This conceptual framework and the research approach will further the past work of examining the degree to which differences in teacher knowledge about reading and writing instruction are associated with student achievement in reading and writing – at least at the level of the primary grades (Bos, Mather, Dickson, Podhajski & Chard, 2001; Cunningham, Perry, Stanovich & Stanovich, 2004; McCutchen, Harry, Cunningham, Cox, Sidman & Covill, 2002; Moats & Foorman, 2003; Phelps & Schilling, 2004).

The Primary Grade Reading & Writing Teacher Knowledge Project has four phases. The first phase was an extensive literature review of research about teacher knowledge and instructional effectiveness, particularly with regard to raising student achievement in schools where achievement has been consistently low. This review resulted in a Reading and Writing Teacher Knowledge Taxonomy. The second phase of the project was to employ modern test development theory to develop the Literacy Instruction Knowledge Scales (LIKS). Based on items in the Taxonomy, the Literacy Instruction Knowledge Scales (LIKS) includes two inter-

related subscales: a pencil-and-paper multiple-choice test of teacher knowledge about reading and a Classroom Observation Rating Scale designed to quantify the quality of classroom instruction in decoding and fluency, comprehension and vocabulary, and writing. The third phase of the project was to establish validity and reliability of the Literacy Instruction Knowledge Scales using generalizability and dependability studies. The final phase of the project will be to establish the predictive validity of the instruments against K-3 standardized reading and writing test scores.

This paper reports on phases one and two of the project.

Methods and Procedures

The Primary Grade Reading & Writing Teacher Knowledge Project was designed to develop a reliable, comprehensive, and practical assessment system to measure primary grade teachers' knowledge of effective, evidence-based reading and writing instruction. The *Literacy Instruction Knowledge Scales (LIKS)* assess primary grade teachers' knowledge of effective, evidence-based reading and writing instruction in two ways: 1) a written survey, and 2) classroom observations. The written survey was designed to assess teacher *pedagogical content knowledge* (Shulman, 1986, 1987; Phelps and Schilling, 2004). The *LIKS* classroom observation subscale was designed to assess teacher *enacted knowledge in classrooms* in the ecologically valid setting of on-going classroom reading and writing instruction (Bronfrenbenner, 1979).

Literacy Instruction Knowledge Scales (LIKS)—Written Survey

In order to construct the *LIKS* written survey, we began with an evidence-based search of the literature base in reading and writing. From the search process, we developed a conceptual framework, called the *Taxonomy of Grade 1-3 Teacher Knowledge of Reading and Writing*

Instruction. This taxonomy was used to develop both the written survey and the classroom observation scale for the *LIKS*.

Developing the taxonomy. To develop an evidence-based framework for describing the knowledge necessary for primary grade teachers to effectively teach children to read and write, we engaged in four distinct, but interlocking, steps intended to triangulate data sources and verify data coding and categorization: 1) an extensive and systematic literature review, 2) classroom observations and interviews with accomplished primary grade teachers who were succeeding in teaching reading and writing to grade 1-3 students in at-risk school settings, 3) teacher focus groups to member check data coding and categorization from the literature review and the classroom observations and interviews, and 4) an audit trail review of the data coding and categorization processes and products.

To provide a systematic framework for conducting the extensive literature review, we relied upon Shulman's (1987) conceptual frame to guide our evolving description of teacher knowledge (Cunningham, Perry, Stanovich & Stanovich, 2004; Phelps & Schilling, 2004). Shulman enumerates seven categories of knowledge for teachers): 1) content knowledge, 2) general pedagogical knowledge, 3) curriculum knowledge, 4) pedagogical content knowledge, 5) knowledge of learners and their characteristics, 6) knowledge of educational contexts, and 7) knowledge of educational ends, purposes, and values. In addition to Shulman's conceptual framework, we were also guided by three related bodies of research, 1) studies of exemplary teaching of reading and writing in the primary grades, 2) measurement studies of reading teacher content and pedagogical knowledge, and 3) evidenced-based preferred practices in reading and writing instruction.

Members of the research team, using the seven categories of Shulman's teacher knowledge conceptual framework, reviewed the past 20 years (1985-2005) of published research in recognized nationally indexed reading and writing research journal article abstracts, major research reports, conference proceedings, or other disseminated syntheses of research on reading and writing instruction, handbooks of research in reading and writing, and major books on teacher knowledge. The literature review included works on reading and writing teacher education; reading and writing research evidence; assessment of reading and writing teacher knowledge; experimental, correlational, and descriptive research investigating the relationships between teacher knowledge of effective reading and writing instruction and children's reading and writing achievement; and literature on effective, evidence-based reading and writing instructional practices (McCardle & Chhabra, 2004; Pressley, Allington, Wharton-McDonald, Block & Morrow, 2001).

Working with other research team members and graduate research assistants, the *LIKS* research team recorded bibliographic references and coded research articles, handbook or book chapters, reports, etc into one of the seven categories of teacher knowledge using an EXCEL spreadsheet. In addition to coding in the different categories, team members identified key reading and writing concepts derived from the literature search.

Because the literature review was iterative in nature, the data collection and coding of concepts became more focused and systematic as specific knowledge domains were interpreted from the data originally coded into Schulman's seven categories (Lincoln & Guba, 1985; Strauss & Corbin, 1990). The *LIKS* research team also gave attention to a search for disconfirming as well as confirming data during the systematic literature review (Erickson, 1986; Shavelson &

Towne, 2002). Data collection and analysis eventually led to theoretical saturation where no new concepts were identified (Glaser & Strauss, 1967; Strauss & Corbin, 1990).

The many concepts derived from the literature search were then placed within larger categories and superordinate concepts, e.g. KWLs and QARs were coded under the superordinate concept of reading instructional materials/curricula/programs. Forty superordinate concepts, called domains of teacher knowledge, were identified between the two areas of reading (20) and writing (20) instruction:

	READING	**WRITING**
1	Learning/Teaching Theory/Models	Learning/Teaching Theory/Models
2	Learning/Teaching Philosophies	Learning/Teaching Philosophies
3	Classroom Environment	Classroom Environment
4	General Reading Instructional Practices	General Writing Instructional Practices
5	Reading Instructional Materials, Curricula, & Programs	Writing Curriculum, Standards, & Programs
6	Texts	Rhetoric and Texts
7	Language/Oral Language	Stages of Writing Development
8	Concepts About Print	Concepts About Writing
9	Phonological/Phonemic Awareness	Modes and Characteristics of Effective Writing Instruction
10	Decoding	Writing Process
11	Spelling Instruction	Spelling
12	Fluency Instruction	Handwriting
13	Vocabulary	Grammar
14	Comprehension	Composition
15	Diversity	Diversity & Writing
16	Motivation/Engagement	Writing Motivation/Engagement
17	Technology and Reading	Technology in writing
18	Reading Assessment	Writing Assessment
19	Family-School Connections	Family-School Connections
20	Teacher Self Awareness	Teacher Self Awareness as Writer or Teacher of Writing

After the *LIKS* research team members completed the literature review, more than 3000 reference citations were cataloged and coded to support the identification of the forty domains of primary grade teacher knowledge about effective, evidence-based reading and writing instruction.

Verifying the taxonomy. Four separate processes were used to verify the taxonomy once it had been established. Classroom observations, teacher focus groups, expert reviewers and an audit member check were used to ensure that the forty domains of knowledge were accurate, inclusive and complete.

First, the *LIKS* research team submitted the forty domains of teacher knowledge to a "reality" check in a series of classroom observations. Two local school districts participating in Reading First, a federally funded school reform project, were purposefully selected to further refine and inform the emerging taxonomy. Four non-Reading First elementary schools were also randomly selected as well from the remaining districts in the area.

Members of the *LIKS* research team observed a set of randomly selected grades 1-3 teachers. Members took field notes of their observations of the reading and writing instruction. Following completion of the classroom observations, research team members compared the items on the field notes to the forty domains of teacher knowledge about reading and writing instruction. In so doing, *LIKS* researchers determined if something observed in teachers' classroom reading and writing instructional behaviors was not accounted for in the taxonomy's forty domains of teacher knowledge. The 15 classroom observations failed to turn up anything in classroom instruction that could not be accounted for within the existing forty domains of teacher knowledge about reading and writing instruction.

Next, we conducted teacher focus groups to triangulate the literature review and teacher observations/interview data in order to increase credibility of the data collection and analysis process. The research team used "member checking" of the data, a process where data are checked with the "members of those stake holding groups from whom the data were originally collected" (Lincoln & Guba, 1985, p. 314). Teacher focus group questions were developed for use in two distinct focus group processes: 1) brainstorming and 2) focused discussion. The forty teacher knowledge domains for teaching reading and writing were used to formulate *enumerative listing questions* for use during brainstorming. Enumerative listing questions required that teachers brainstorm items they believe should be identified as representing effective, evidence-based teacher knowledge of reading and writing instruction.

Teachers were then provided with ½ of the reading or writing sections of the *Taxonomy*. Using a coding example, the teachers indicated whether they agreed, disagreed, or were uncertain or had a question that a semantic or meaningful relationship existed between each coded data item and the category or domain to which the item was assigned within their ½ of the *Taxonomy*. This distribution allowed at least three to four raters on each portion of the two halves of the *Taxonomy*. Teachers were asked to add any additional items or categories that, in their judgment, should be included on the listing of teacher knowledge domains identified during the classroom observations, teacher interviews, and the literature review.

High frequency responses to the enumerative listing questions were compared to the forty teacher knowledge domains for teaching reading and writing identified during the classroom observations, teacher interviews, and the literature review were used and judged as: (a) confirming these categories and items, (b) new items that should be added to the categories or coded data list, or (c) irrelevant items which should be eliminated from consideration. Low

frequency or single incidences of data were viewed as discrepant cases and analyzed to determine if existing categories or coded data items need to be modified, refined, or elaborated (Goetz & LeCompte, 1984: LeCompte & Preissle, 1993).

A third method of verifying the *Taxonomy* was through expert review. Prior to finalizing the *Taxonomy* for dissemination, we sent it out for review by two nationally prominent reading/writing experts and two nationally board-certified primary grade teachers. These external expert reviews generally confirmed the forty teacher knowledge domains for primary grade reading and writing instruction. Very few of the expert reviewers suggested deletions, additions, or modifications, and none of the suggestions was found to converge among two or more of the expert reviewers.

Finally, an audit trail review of the data collection and coding verification process was conducted using collected, analyzed and coded data drawn from the triangulation processes of the literature review, classroom observation, and focus group meetings. The purpose of an audit trail review was for outside "auditors" to verify the assignment of category, count, and coding of data items into categories or domains from the listing produced in the activities of literature review, classroom observations and interview, and the focus group meetings. Two competent individuals trained in qualitative research methodologies were contacted to serve as outside data auditors (Lincoln & Guba, 1985; Osterlind, 1989). The audit process required two days which allowed one day for browsing through the data, and one day to carry out the audit and prepare the report found in the auditors' letter of attestation which is on file with the *LIKS* research team. (The finalized *Taxonomy* with the supporting references coded by domain are web-accessible at www.coe.usu.edu/ecc).

Developing the multiple-choice items. Based on the *Taxonomy* the research team began the development of the written knowledge survey of the *LIKS* that included: (a) writing a statement of purpose, (b) developing domain specifications to guide the writing of survey items, (c) writing instrument specifications, and (d) developing teacher knowledge survey items.

First, a clearly defined purpose statement was developed to describe the population of users for which the *LIKS* was intended, the purposes for which the resulting scores were intended to be used, and the constraints under which the *LIKS* should be used (i.e., grade levels, time required and preferred conditions for response or observations, one-time assessment or repeated use to measure growth, etc., Crocker & Algina, 1986; Ebel & Fiesbie, 1991; Osterlind, 1989; Thorndike, 1982). This statement is found below:

The purpose of the Literacy Instructional Knowledge Scales (LIKS) is to assess grade 1-3 teacher knowledge of effective, evidence-based reading and writing instruction. The scores derived from the LIKS may be used to determine professional development needs within school districts or to examine the substance of teacher education programs related to grade 1-3 teachers' knowledge of effective, evidence-based reading and writing instruction. The written survey subscale of the Literacy Instructional Knowledge Scales (LIKS) is intended to be group administered under secure conditions and requires approximately 90 minutes for completion.

Second, the research team developed domain specifications (i.e., format, categories, and descriptors) as an organizational scheme to guide the writing of individual items. Writing a teacher knowledge survey that would address all forty domains listed in the *Taxonomy* in an adequate and reliable manner was deemed neither useful nor practical for several reasons. First, to write a survey addressing forty separate domains of teacher knowledge about reading and writing instruction would require the development of an exceedingly large item pool, ostensibly

requiring upwards of 800 individual items. Second, a reliable written test covering forty separate domains of teacher knowledge about reading and writing instruction would be impractical to administer in a reasonable timeframe as well as requiring well over 3000 responders for item analysis. Also, the research team determined that not all forty domains of teacher knowledge about reading and writing instruction—for example, knowledge of philosophies, theories—were likely to have an impact upon teacher enacted knowledge in the classroom setting and therefore would also be unlikely to influence grade 1-3 student achievement in reading or writing.

Thus, it was decided that, for the sake of test construction, items developed would reflect teacher knowledge likely to impact teacher enacted knowledge and student achievement. Therefore, the forty domains of teacher knowledge about reading and writing instruction were collapsed into four "super" domains for test development purposes. These four superdomains were: a) classroom climate and management, b) decoding, c) comprehension, and d) writing.

Members of the *LIKS* research team were of the opinion, supported by Shulman's (1986, 1987) conceptual framework of teacher knowledge and related research in exemplary classrooms and beyond, that classroom climate and management were necessary but insufficient conditions for providing high quality evidence-based literacy instruction (Pressley, Allington, Wharton-McDonald, Block & Morrow, 2001; Taylor, Pearson, Clark & Walpole, 1999; Taylor, Pearson, Peterson & Rodriguez, 2005). We wanted to test empirically the findings of past research and the assertion of researchers and classroom practitioners that classroom climate and management uphold and/or undergird high quality primary grade reading and writing instruction. The *classroom climate and management* superdomain initially focused on five domains in the *Taxonomy*: 1) access to literacy materials, 2) spatial arrangement of the classroom to support

literacy instruction, 3) differentiated instruction, 4) rules, routines, and procedures, and 5) interaction and engagement. Later in the development process, a sixth domain dealing with instructional pacing was added.

The *decoding* superdomain initially focused written survey items on five domains found in the *Taxonomy*: 1) alphabetics (letter name/sound knowledge), 2) blending (letters and syllables), 3) spelling/segmenting, 4) sight words, and 5) fluency practice. The *comprehension* superdomain focused written survey item writing on five domains in the *Taxonomy*: 1) asking questions, 2) activating background knowledge, 3) comprehension strategy instruction, 4) word meanings, and 5) comprehension practice. The final superdomain, writing instruction, initially focused written survey item writing on four domains in the *Taxonomy*: 1) writing practice 2) modeling writing process, 3) written language study, and 4) identifying writing task, purpose.

Next, instrument specifications defined the item format, characteristics of item descriptors, number of items, directions for responding to each item, and procedures for scoring (Crocker & Algina, 1986; Ebel & Fiesbie, 1991; Osterlind, 1989; Thorndike, 1982). The item format included, (a) a code for each of the four teacher knowledge domains using an alphabetic abbreviation representing the teacher knowledge domain and a numeric indicator to represent the item number, and (b) a generalized multiple choice response format. We determined that using a multiple-choice format for developing the written survey subscale items of the *LIKS* was supported by previous findings that multiple-choice items have been found to effectively and efficiently measure teacher pedagogical content knowledge as compared to different and other more elaborate item writing approaches such as situated judgments, constructed responses, high fidelity simulations, Q sorts or checklist items (Phelps & Schilling, 2004).

The *LIKS* research team determined that written items should tap teacher knowledge from, 1) the four "super" domains developed from the *Taxonomy,* 2) enacted knowledge as would be "observed" in classroom reading and writing instruction, and 3) assessment knowledge and its connection to teaching decisions in reading and writing. The number of items to be developed for each of the four "super" domains was slightly more heavily weighted toward decoding and comprehension in reading than classroom climate and management and writing based upon the amount and quality of available research evidence available to support item development. The initial item pool numbers, types of teacher knowledge to be assessed, and the domains of teacher knowledge for the *LIKS* written survey subscale pilot item pool are shown in the table of specifications below (Thorndike, 1982):

Climate and Classroom Management 30-40 questions 20%	Decoding embedding assessment but separable 45-60 items 30%	Comprehension embedding assessment but separable 45-60 items 30%	Writing embedding assessment but separable 30-40 items 20%	Teacher Knowledge to be Assessed
20 from obs. 10 taxonomy	30 from obs. 20 taxonomy	30 from obs. 20 taxonomy	20 from obs. 10 taxonomy	(EKC) (PCK)
10 assessment	10 assessment	10 assessment	10 assessment	Assessment

The writing of the initial *LIKS* subscale item pool totaling a minimum of 200 items extended over a period of eight months of revisions and review. The 200 initial items in the *LIKS* written subscale item pool were reduced to a total of 150 items for the pilot written survey teacher knowledge subscale of the *LIKS*. Draft example items for each of the four domain specifications

– classroom climate and management, decoding, comprehension, and writing, for the *LIKS* written survey subscale item pool – are shown in Appendix A.

The items for the *LIKS* written survey subscale pilot version went through multiple rounds of internal revisions, including reviews by research team members and experts as well. These review processes included individual, paired, and whole group item-by-item reviews and discussion. The purpose of each round of item review and revision was to refine relevance, focus, and content of items, improve item formatting and style, and converge on correct responses. Each round of review and revision included an evaluation of the wording (i.e., accuracy, readability, and clarity), discreteness and distribution of items, stylistic uniformity, quality and function of distracters, type of teacher knowledge assessed, and consensus and convergence upon correct responses.

Literacy Instruction Knowledge Scales (LIKS)—Classroom Observation Scale

We based the development of the second subscale, the classroom observation subscale of the *LIKS*, on informed professional judgment and the elements of effective, evidence-based reading and writing instruction as listed in the *Taxonomy of Grade 1-3 Teacher Knowledge of Reading and Writing Instruction.* The development of the classroom observation subscale made use of the same *statement of purpose* and *domain specifications* – classroom climate and management, decoding, comprehension, and writing— as the written survey subscale of the *LIKS*. However, the subscale specification and items for the classroom observation subscale of the *LIKS* differed significantly from the written survey subscale.

Instrument specifications. The item format for the *LIKS* classroom observation subscale was a set of rubrics based on a seven-point Likert type rating scale with descriptors under points 1, 3, 5, and 7 on the scale. Within the classroom climate and management domain for the

classroom observation subscale, six sections were identified: 1) access to literacy materials, 2) spatial arrangement of classroom to support literacy instruction, 3) differentiated instruction, 4) rules, routines and procedures, 5) interaction and engagement, and 6) pacing. Within the decoding domain for the classroom observation subscale, four sections were identified: 1) fluency, 2) phonemic awareness and phonics, 3) application of decoding to reading text, and 4) regular and irregular high-frequency words. Within the comprehension domain for the classroom observation subscale, five sections were identified: 1) comprehension instruction before reading, 2) comprehension instruction during reading, 3) comprehension instruction after reading, 4) comprehension strategy instruction, and 5) vocabulary instruction. Within the writing domain for the classroom observation subscale, four sections were identified: 1) writing lessons, 2) guided writing practice, 3) writing tasks, and 4) audience and response to student writing.

Each section of the classroom observation subscale of the *LIKS* contained at least two items that could be rated from 1-7. For example, in the section on classroom climate and management, *Access to Literacy Materials*, four items were identified, 1) quantity of reading materials, 2) quantity of displayed print in the classroom, 3) writing supplies, and 4) organization of reading and writing materials for storage, use, and display. Each item contained a written rubric for the rating scale, 1-7. For example, under *Access to Literacy Materials*, 1) quantity of reading materials, the rubrics were as follows: 1= no additional reading materials, 3= 1-100 books in classroom library, no magazines, dictionaries, school textbooks, 5= 101-299 books in classroom library, 1-2 magazine titles, dictionary, encyclopedias, school textbooks, and 7= 300 books of differing levels and genres in classroom library, 3+ magazine titles, variety of reference materials, school textbooks. Observers were required to select a rubric and corresponding number from 1-7 that most closely represented what they saw in the classroom. The even

numbers, 2, 4 and 6 were used when observers wanted a rating in between the odd numbered rubrics.

Scale development. The *LIKS* research team repeatedly met and reviewed the development of the sections and items to be included in the classroom observation subscale. The team members repeatedly asked the question, *"What could and should we expect to see on any given day as a part of effective, evidence-based reading or writing instruction?"* This question, along with the evidence-base already described in the *Taxonomy,* guided the development of the *LIKS* classroom observation subscale. Draft example items for each of the four domain specifications for the *LIKS* classroom observation subscale item pool can be found in Appendix B.

Linking the LIKS classroom observation scale to the LIKS written survey. A critical step in the *LIKS* test construction process was the linkage between the *LIKS* classroom observation scale back to the *LIKS* written survey. In order to link the two subscales together, the *LIKS* research team asked the question, *"What would teachers need to know in order to achieve a 7 on each section of the LIKS classroom observation subscale?"* The research team revisited all the items on the written survey to identify those items that linked directly to the *LIKS* classroom observation scale. For example, for the *Access to Literacy Materials* section, quantity of reading materials, the team searched the item pool for specific questions that teachers would need to know the answers to in order for them to enact that knowledge in the classroom. So, for example, for the quantity of reading materials item, if teachers did not understand the importance of student access to multiple kinds of reading materials, they would be unlikely to have such materials available in their classrooms.

For each section of the *LIKS* classroom observation subscale, the research team identified at least five written survey items that would link back to the observation scale. If there were not a sufficient number of items, then the research team wrote new items to add to the item pool.

Classroom observation item review procedures. The *LIKS* classroom observation subscale item pool went through a variety of review iterations ranging from five revision cycles to over twenty revision cycles stretching over nearly a year of development. The first iterative review of the *LIKS* written survey items was conducted by using the classroom observation subscale to observe grades 1-3 classrooms. During the observations, pairs of raters from the research team scripted teacher behaviors and instruction using field notes during the entire language arts instructional block in grade 1-3 classrooms. After scripting, the two raters independently determined which of the sections of the *LIKS* classroom observation subscale to rate and then rated each of the items within the sections. Early informal assessments of inter-rater agreement ranged from 80 to 100% agreement among pairs of raters for the initial try out. Following this initial try-out, several problems with the descriptors were identified and corrected.

The next iterative review for the classroom observation subscale utilized videotaped reading lessons. These lessons had been videotaped for other purposes and were used for the purpose of evaluating the classroom observation scale. After viewing the video together and scripting and rating the events independently, research team members discussed the ratings for each item. Ambiguous descriptors were discussed using “think alouds” after each item. Again, major revisions to item descriptors were made specifically dealing with the scaling responses - what produced a 1, 3, 5 or 7 rating.

Following revisions suggested by the *LIKS* research team to the item descriptors, the revised items were tried out informally with graduate students, reading coaches, and school

administrators in live classroom observations. The raters scripted and rated their observations independently. After doing so, research team members collected the rating scales and computed percentages of agreement or correlations among the raters. Again, even among untrained users of the *LIKS* classroom observation subscale, agreements were very high.

Sub groups were subsequently assigned to watch two existing DVD based reading and writing classroom language arts instructional blocks. They viewed these DVD lesson blocks and scripted and rated each independently, discussed disagreements or continuing problems with the instrument, and shared their findings with the entire research team. Following this revision, the pilot version of the LIKS classroom observation subscale was readied for use in a full scale generalizability and multi-facets Rasch analysis using a nested design to systematically investigate sources of error variance and reliability.

Procedures for Pilot Testing the LIKS—Written Survey

Permission was granted from nine school districts within the state to collect the data required for the research and development of the *LIKS* written survey. Elementary schools in the districts were randomly selected for recruiting grade 1-3 teachers to complete the pilot version. A school district research coordinator was identified in each of the participating districts to coordinate the administration of the *LIKS* in each district.

LIKS research team members created training procedures and materials, including sufficient numbers of *LIKS* written survey subscale pilot instruments, sufficient numbers of computer bubble sheets for participants to register responses, and a recruitment script for talking with elementary school principals and for recruiting elementary school teachers to complete the *LIKS* written survey. All materials were placed into a prepaid, return postage box for training.

Training of the district level research coordinators was conducted in the district offices of a centrally located school district for three hours by three members of the research team during which the *LIKS* research project was described, school district research coordinator role statements were reviewed, school district research coordinator contracts were signed and school district research coordinator questions were answered. During the data collection period, any questions that arose from the school district research coordinators were answered by one of the PIs on the *LIKS* research project via e-mail or phone.

Procedures for Pilot Testing the LIKS—Classroom Observation Scale

The *LIKS* research team videotaped 20 primary grade teachers in two selected schools districts in Utah. Seven below average teachers, seven average teachers, and six above average teachers were purposefully selected from the two school districts. Once videotaping was completed, the research team met to view the first videotaped teacher as a group to work through the process of viewing the teacher and using the *LIKS* classroom observation subscale. Next, team members viewed teachers 2 and 3 in subgroups. The team met again to compare their independent ratings on these teachers. Finally, the measurement expert worked out the assignment of research team members to view and rate the remaining teachers, DVDs # 4-20, independently as part of the generalizability study of the *LIKS* classroom observation subscale. Analysis and results for the pilot testing of the *LIKS –Classroom Observation Subscale Pilot Version* is not yet complete.

Results

The initial version of the knowledge test included 150 items in four subdomains. The questions within each subdomain were tentatively classified into two types of knowledge:

(a) pedagogical content knowledge, and (b) knowledge of research-based principles for teaching reading and writing. Table 1 displays the distribution of the 150 items classified by subdomain and type of knowledge. Each question in the test is represented by its item number in this table.

The 150-item test was administered on a trial basis to 504 practicing teachers. The main purposes of this tryout were to collect item response data that could be used (a) to assess how well the various items and distracters functioned, and (b) to make informed decisions about which items should be revised, deleted, or retained in order to improve the test.

The teachers recorded their answers on a scannable answer sheet. The responses of 2 of the 504 teachers were excluded from further analysis because one had omitted more than 25 items and the other had made a mistake somewhere on the answer sheet and marked 151 responses instead of 150. The responses of the remaining 502 teachers were analyzed using three different approaches:

1. classical item analysis statistics including distracter analysis (using *SPSS*)
2. the 1-PL Rasch model (Using *Winsteps*)
3. the 2-PL IRT model (using *BILOG-MG*)

The results of the three analyses complemented each other and led to the same general conclusions. For the purposes of this paper we have chosen to focus on reporting the results of the classical item statistics and distracter analysis.

The reliability of total scores obtained from the complete 150-item test estimated by Cronbach's alpha coefficient was .816. Similarly, the Rasch reliability estimate obtained from *Winsteps* was .81. Since a major purpose of the test is to estimate the knowledge of each teacher-examinee on each subscale, the reliability of the subscale scores is more important than the reliability of the total scores. Ideally, one would expect scores from a test with 150 items to have

a higher reliability than .81. The heterogeneous nature of the 150 items across the four subdomains and the two different types of knowledge may be one reason for the .81 reliability coefficient.

However, the wide range of item difficulties appears to be another reason for this low reliability estimate. The size of the correlation between any two items is limited by the shapes of the distributions of the two items. A maximum value of 1.0 for the correlation between two variables is possible only when the distributions of the two variables have the same identical shape. Of course, the degree to which two variables are correlated depends on the degree to which scores on one variable covary with scores on the other, but if the distributions of the two variables differ in the shape, this difference constrains the maximum possible value of the correlation coefficient. This constraint is especially problematic for variables that are dichotomously scored, because if one item in the test is extremely easy and another item is quite hard, then the maximum possible correlation will be substantially less than 1.0 (Carroll, 1961; Glass & Hopkins, 1996; Guilford & Fruchter, 1973; Nunnally & Bernstein, 1994).

The easiest question on the 150-item test was Item 12 which was answered correctly by 498 (99.2%) of the teachers. Conversely, the most difficult question was Item 97 which was answered correctly by only 43 (8.6%) of the 502 teachers answered this item correct. Figure 1 displays the distributions of these two items and reveals that the shapes of the two distributions are very dissimilar. In fact, they are almost opposite. One distribution is markedly skewed to the left, while the other is markedly skewed to the right. This disparity in shape severely limits the maximum possible value of the correlation coefficient. In fact the maximum possible correlation between dichotomous variables that are this disparate in difficulty is only .03. [This value was computed using the formula for the maximum value of a phi coefficient provided in Carroll

(1961) and also in Guilford and Fruchter, (1973)]. Given this difference in the shapes of the two distributions, it would not be surprising to find that the actual correlation is negative.

The size of Cronbach's alpha coefficient varies as a function of the number of items in a test and size of the average inter-item correlation. As the size of the average inter-item correlation increases, the size of the resulting alpha coefficient increases although the relationship is not linear. The inter-correlations for the 150-item ranged from -.215 to .365 with a mean of .030. This low value is a direct result of the wide disparity in difficulty of the various items.

Eleven of the 150 items had difficulty indices less than .20 (meaning that less than 20% of the teachers answered these items correct). On the other hand, 26 items were extremely easy for this group of teachers (more than 90 percent of the 502 teachers answered these items correct) and another 24 items had item difficulties between .80 and .89. This means that one-third of the test items (26 + 24 = 50 out of 150) were very easy. Consequently, a large number of the inter-item correlations were constrained in magnitude by the disparities in the shapes of their distributions. The solution to this lack of inter-item consistency is to perform a distracter analysis and then either revise the extremely easy and extremely hard items so that they have less extreme difficulty indices or delete them. Either of these actions will help to increase the average inter-item correlation and coefficient alpha. It is not necessary to use item response theory to reach this conclusion, although adjusted item-to-total correlations from SPSS and the point biserial correlations from the 1-PL Rasch analyses confirm this conclusion.

As indicated previously, we are more concerned about the reliability of the various subscale scores than we are about the reliability of the total scores obtained from the test because the various subscales are presumed to assess different dimensions or factors. However, before

reporting the subscale results it is necessary to explain one other change that we decided to implement. We decided to delete the Classroom Management subscale and to completely remove the 30 items on this subscale. (This decision was prompted earlier by concerns about the practicality of obtaining reliable and valid measures of teachers' classroom management practices to correlate with their scores on the knowledge test).

In addition to deleting the 30 Classroom Management items, we removed 10 other items based on the initial item analysis results. These 10 included 1 from the Decoding scale, 4 from the comprehension scale, and 5 from the writing scale. We then re-computed the item analysis statistics based on the remaining 110 items. Table 2 presents summary statistics for each of the six remaining subscales.

Note that the alpha reliability coefficients are all low. Visual inspection of the three columns reported under the Item Difficulty spanner reveals that there is still a wide range between the minimum and maximum item difficulty levels that were observed. These disparities reduce the inter-item correlations. There are far too many inter-item correlations to report individually so they are summarized in Table 2. Note that the average inter-item correlation for each of the six subscales is less than .06. Again, the solution to this problem is to revise the easy items and make them more difficult (better functioning distracters) and to carefully examine the difficult items and make whatever revisions are warranted. The distracter analysis data as well as the point biserial correlation coefficients will be used to inform this process.

Preliminary distracter analysis indicates that most of the items are likely to be useful and simply need to be fine tuned by revising one or two weak or nonfunctioning distracters. This conclusion is confirmed by the results of a preliminary confirmatory factor analysis,

although given the observed lack of reliability plus the fact that many of the items need further revision we consider the results of the factor analysis to be preliminary and tentative. The results support the existence of three discrete factors with reasonably good fit and at least 15 good items for subscale. These three factors are reasonably similar to our three major subdomains of decoding, comprehension, and writing. We plan to do further factor analyses after we improve the items.

At first glance, the analysis of our trial test results appear to be disappointing. These results have certainly humbled us and helped us to face reality. On the other hand, we believe that there is a silver lining to this cloud and the initial results are quite promising. We now know that we have (a) some good items, (b) many others that have potential but need to be revised and fine-tuned, and others (c) that appear to lack potential and should be deleted. Most importantly, we now have empirical evidence to help us know which items should be classified into each of these three categories.

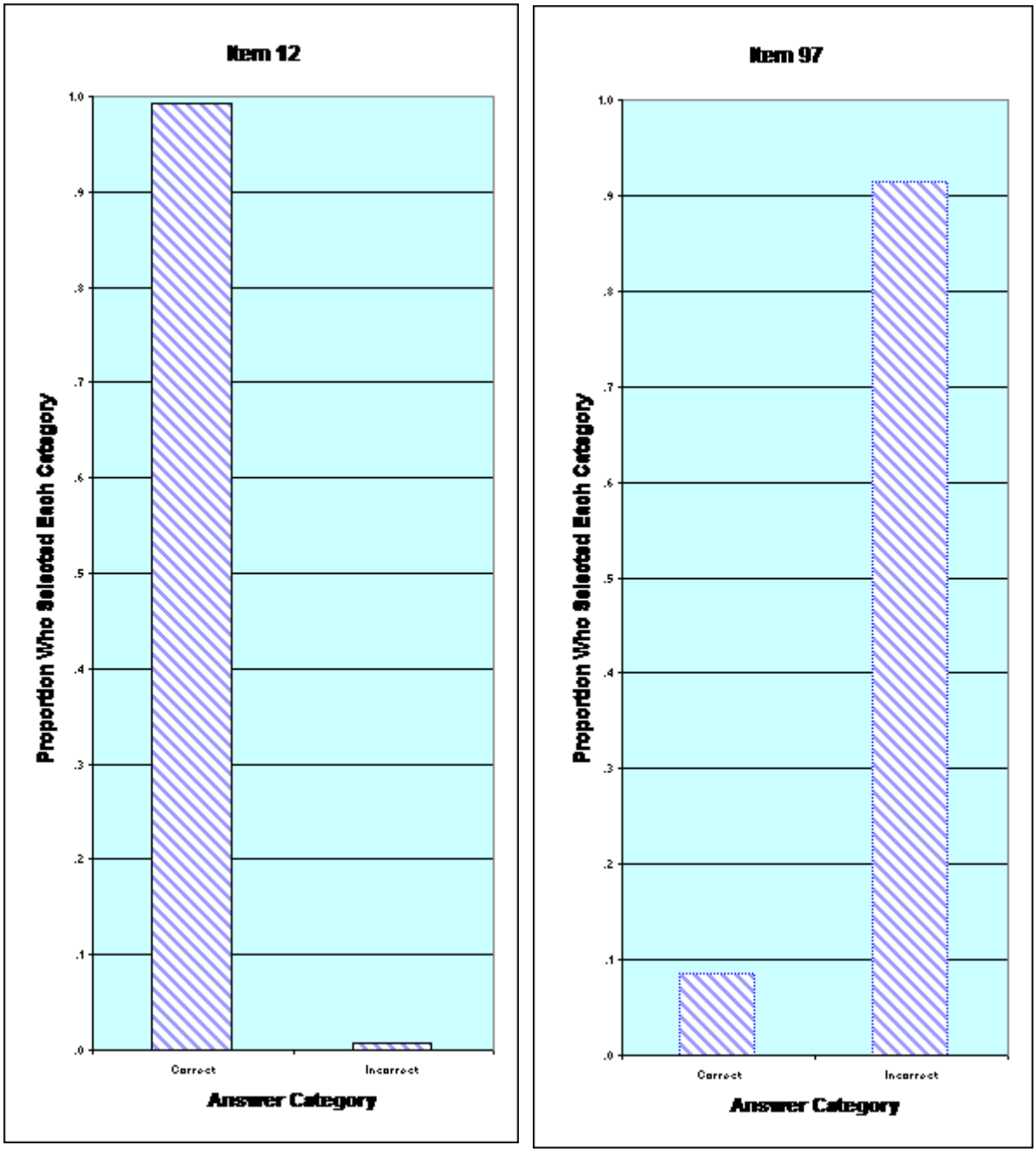


Figure 1. Distribution of Responses to Items 12 and 97

Table 1.

Classification of the 150 Questions in the Teacher Knowledge Test by Subdomain and Type of Knowledge

Type of Knowledge	Subdomain							
	Decoding		Comprehension		Writing		Classroom Managmnt	
Pedagogical Content Knowledge	4	92	17	68	28	134	13	82
	7	93	20	72	35	137	19	109
	11	97	33	75	47	139	36	110
	25	99	49	105	48	140	39	113
	46	103	50	108	64	141	52	114
	54	104	51	122	65	142	71	116
	63	106	55	147	87	143		
	69	107	56	148	128	144		
	73	111	57	149	130	145		
	77	117			131	146		
	80	118			132			
	81	120						
	85	125						
	88	126						
	89	127						
Knowledge of Research-based Principles	1	61	16	58	100		2	67
	5	74	18	59	133		3	68
	6	90	22	76	136		8	70
	14	95	27	78	138		9	101
	26	96	29	79			21	102
	32	98	30	84			23	112
	60	124	31	86			38	115
			41	91			43	119
			42	94			62	121
			44	123				
			45	150				
			53					

Table 2. Summary of Item Statistics by Subscale

Type of Knowledge	Subscale	Number of Items	Item Difficulty Levels			Inter-item Correlations			Corrected Item-total Correlations		Coefficient alpha
			Minimum	Maximum	Mean	Minimum	Maximum	Mean	Minimum	Maximum	
Content Knowledge	Decoding	30	.086	.974	.543	-.136	.287	.049	-.049	.377	.615
	Comprehension	18	.112	.964	.642	-.105	.324	.041	.029	.194	.426
	Writing	21	.213	.970	.702	-.104	.334	.052	.032	.315	.519
Knowledge of Research-based Principles	Decoding	14	.131	.972	.632	-.212	.227	.027	-.058	.204	.280
	Comprehension	23	.193	.990	.728	-.102	.261	.042	.019	.280	.494
	Writing	4	.149	.930	.554	-.035	.095	.020	-.007	.055	.060

Selected References

Anderson, L. M., Evertson, C. M., Brophy, J. E. (1979). An experimental study of effective teaching in first-grade reading groups. *The Elementary School Journal, 79*(4), 193-223.

Arlin, M. (1979). Teacher transitions can disrupt time flow in classrooms. *American Educational Research Journal, 16*, 42-56.

Armbruster, B., Lehr, F., & Osborn, J. (2001). *Put reading first: The research building blocks for teaching children to read, K-3*. Washington, DC: National Institute for Literacy.

Asselin, M.M. (1997). Teacher's knowledge of reading instruction: A description and investigation of effects. *Alberta Journal of Educational Research, 43*(4), 235-52.

Auten, A. (1985). Small group instruction: Effective classroom management? *Journal of Reading, 28*(5), 460-462.

Barr, R., Kamil, M. L., Mosenthal, P., & Pearson, P. D. (1991). *Handbook of reading research, volume II.* New York: Longman .

Baumann, J.F., Hoffman, J.V., Duffy-Hester, A., & Ro, J.M. (2000). The first R yesterday and today: U.S. elementary reading instruction practices reported by teachers and administrators. *Reading Research Quarterly, 35(3),* 338-377.

Bos, C., Mather, N., Dickson, S., Podhajski, B., & Chard, D. (2001). Perceptions and knowledge of preservice and inserice educators about early reading instruction. *Annuals of Dyslexia, 51*, 97-120.

Bronfenbrenner, U. (1979). *The ecology of human development*. Cambridge: Harvard University Press.

Brophy, J. (1981). Teacher praise: A functional analysis. *Review of Educational Research, 51*(1), 5-32.

Brophy, J. (1983). Research on the self-fulfilling prophecy and teacher expectations. *Journal of Educational Psychology, 75*(5), 631-661.

Cameron, J. & Pierce, W. D. (1994). Reinforcement, reward, and intrinsic motivation: A meta-analysis. *Review of Educational Research, 64*(3), 363-423.

Carroll, J.B. (1961). The nature of the data, or how to choose a correlation coefficient. *Psychometrika, 26,* 347-372.

Cochran-Smith, M. & Zeichner, K.M. (2005). *Studying Teacher Education: The Report of the AERA Panel on Research and Teacher Education*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory.* New York: Harcourt Brace Jovanovich.

Cunningham, A.E., Perry, K.E., Stanovich, K.E., & Stanovich, P.J. (2004). Disciplinary knowledge of K-3 teachers and their knowledge calibration in the domain of early literacy. *Annals of Dyslexia, 54,* 139-167.

Darling-Hammond, L. (2006). *Powerful teacher education: Lessons from exemplary programs*. San Francisco: Jossey-Bass.

Darling-Hammond, L., Bransford, J., & LePage, P. (Eds). (2007). *Preparing Teachers for a Changing World: What Teachers Should Learn and Be Able to Do*. San Francisco: Jossey-Bass.

Durkin, D. (1979). What classroom observations reveal about reading comprehension instruction. *Reading Research Quarterly, 14*(4), 481-533.

Ebel, R. L., & Fiesbie, D. A. (1991). *Essentials of educational measurement* (5th ed.). Englewood Cliffs, NJ: Prentice Hall.

Erickson, F. (1986). *Qualitative methods in research on teaching*. In Wittrock, M. C. (Ed.), Handbook of research on teaching (3rd ed.) (pp. 119-161). New York: Macmillan.

Farstrup, A.E., & Samuels, S.J. (Eds.). (2002). *What research has to say about reading instruction* (3rd ed.). Newark, DE: International Reading Association.

Fenstermacher, G.D. (1994). The knower and the known: The nature of knowledge in research on teaching. *Review of Research on Teaching*, *20,* 1-54.

Fink, J. & Siedentop, D. (1989). The development of routines, rules, and expectations at the start of the school year. *Journal of Teaching in Physical Education*, *8*(3), 198-212.

Flower, L. (1994). *Making thinking visible: Writing, collaborative planning, and classroom inquiry.* Urbana, IL: National Council of Teachers of English.

Flower, L., & Hayes, J. R. (1980). The dynamics of composing: Making plans and juggling constraints, p. 31-50. . In L.W. Gregg & E. R. Steinberg (Eds.), *Cognitive processes in writing.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Gitomer, D. H., Latham, A. S., & Ziomek, R. (1999). The academic quality of prospective teachers: The impact of admissions and licensure testing. Retrieved from www.etc.org/praxis.

Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research.* Chicago: Aldine.

Glass, G.V., & Hopkins, K.D. (1996). *Statistical methods in education and psychology* (3rd ed.). Boston: Allyn and Bacon.

Goetz, J. P., & LeCompte, M. D. (1984). *Ethnography and qualitative design in educational research.* New York: Harcourt Brace Jovanovich.

Goldhaber, D.D. & Brewer, D. J. (1998). When should we reward degrees for teachers? *Phi Delta Kappan, 80,* 2, 134-138.

Good, T.L., & Grouws, D. (1977). Teaching effects: A process-product study in fourth grade mathematics classrooms. *Journal of Teacher Education, 28*(3), 49-54.

Greenwald, R., Hedges, L., & Laine, R. (1996). The effects of school resources on student achievement. *Review of Educational Research, 66,* 361-396.

Guilford, J.P., & Fruchter, B. (1973). *Fundamental statistics in psychology and education* (5th ed.). New York: McGraw-Hill.

Hayes, J. R. & Flower, L. (1980). Identifying the organization of writing processes, p. 3-30. In L.W. Gregg & E. R. Steinberg (Eds.), *Cognitive processes in writing.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Hoffman, J.V., & Pearson, P.D. (2000). Reading teacher education in the next millennium: What your grandmother's teacher didn't know that your granddaughter's teacher should. *Reading Research Quarterly, 35,* 28-44.

Husu, J. (2003). *"Tell me about your work and I will tell you about your knowledge": Constructing teacher knowledge based on pedagogical practices*. Paper presented at the Biennial Conference of the International Study Association on Teachers and Teaching. Leiden, The Netherlands, June 27-July 1, 2003.

International Reading Association (2007). *Standards for Reading Professionals*. Retrieved from http://www.reading.org/.

Kamil, M.L., Mosenthal, P.B., Pearson, P.D., & Barr, R. (Eds.). (2000). *Handbook of Reading Research (Vol. 3).* Mahwah, NJ: Erlbaum.

Kolis, M. & Dunlap, W.P. (2004). The knowledge of teaching: The K3P3 Model.

LeCompte, M. D., & Preissle, J. (1993). *Ethnography and qualitative design in educational research* (2nd ed.). New York: Academic Press.

Levine, A. (2006). *Educating School Teachers*. Washington, DC: Education Schools Project.

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Beverly Hills, CA: Sage.

McCardle, P., & Chhabra, V. (2004). *The voice of evidence in reading research.* Baltimore, MD: Paul H. Brookes Publishing Co.

McCutchen, D., Abbott, R.D., Green, L.B., Beretvas, S. N., Cox, S., Potter, N.S., et al. (2002). Beginning literacy: links among teacher knowledge, teacher practice, and student learning. *Journal of Learning Disabilities 35*(3), 69-86.

McCutchen, D., Harry, D, R,, Cunningham, A. E., Cox, S., Sidman, S., & Covill, A. E. (2002). Reading teachers' knowledge of children's literature and English phonology. *Annals of Dyslexia, 52*, 207-228.

Meijer, P.C., Verloop, N., & Beijaard, D. (2001). Similarities and differences in teachers' practical knowledge about teaching reading comprehension. *Journal of Educational Research, 94*(3), 171-184.

Miller, J., McKenna, J., & McKenna, B. (1998). A comparison of alternatively and traditional prepared teachers. *Journal of Teacher Education, 49,* 165-176.

Mitchell, K. J., Robinson, D. Z., Plake, B. S., & Knowles, K. T. (2001). *Testing teacher candidates: The role of licensure tests in improving teacher quality*. Washington, DC: National Research Council.

Moats, L. (2000). *Speech to print: Language essentials for teachers*. Baltimore: Paul H. Brooks.

Moats, L.C., & Foorman, B. R. (2003). Measuring teachers' content knowledge of language and reading. *Annals of Dyslexia, 53*, 23-45.

Mosenthal, J., Lipson, M., Torncello, S., Russ, B., & Mekkelsen, J. (2004). Contexts and practices of six schools successful in obtaining reading achievement. *The Elementary School Journal, 104*(5), 343-367.

National Assessment of Educational Progress. (2005). Washington, DC: Institute for Education Sciences.

National Institute of Child Health and Human Development. (2000). Report of the National Reading Panel. T*eaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction* (NIH Publication No. 00-4769). Washington, DC: U.S. Government Printing Office.

National Reading Panel (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading its implications for reading instruction.* Washington, DC: National Institute of Child Health and Human Development.

Nunnally, J.C., & Bernstein, I.H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.

Osterlind, S. J. (1989). *Constructing test items*. Boston: Kluwer.

Paris, S., Lipson, M., & Wixson, K. (1983). Becoming a strategic reader. *Contemporary Educational Psychology*, *8*, 293-316.

Pearson, P.D. Barr, R., Kamil, M.L., & Mosenthal, P. (1984). *Handbook of reading research.* New York: Longman.

Phelps, G. & Schilling, S. (2004). Developing measures of content knowledge for teaching reading. *The Elementary School Journal*, *105*, 1, 32-48.

Pressley, M., Allington, R. L., Wharton-McDonald, R., Block, C. C., & Lesley, M. M. (2001). *Learning to read.* New York, NY: The Guilford Press.

Pressley, M., Wharton-McDonald, R., Allington, R., Block, C.C., Morrow, L. M., Tracey, D., Baker, K., Brooks, G., Cronin, J., Nelson, E., & Woo, D. (2001). The nature of literacy instruction in ten grade-4/5 classrooms in upstate New York. *Scientific Studies in Reading, 2*, 159-191.

Rosenshine, B. (1983). Teaching functions in instructional programs. *The Elementary School Journal, 83*(4), 335-351.

Ruddell, R.B., Ruddell, M.R., & Singer, H. (1994). *Theoretical models and processes of reading* (4th ed.). Newark, DE: International Reading Association.

Sanders, W. L. & Horn, S. P. The Tennessee value-added assessment system (TVAAS): Mixed-model methodology in educational assessment. *Journal of Personnel Evaluation in Education, 8* (3), p. 299-311.

Shavelson, R., & Towne, L. (Eds.), (2002). *Scientific research in education.* Washington, D.C.: National Academies Press.

Shulman, L. S., & Grossman, P. (1988). *Knowledge growth in teaching: a final report to the Spencer Foundation*. Stanford, CA: Stanford University.

Shulman, L.S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher, 15*, 4-14.

Shulman, L.S. (1987). Knowledge and teaching: Foundations for the new reform. *Harvard Educational Review, 57*, 1-22.

Snow, C. E., Burns, M. S., & Griffin, P. (1998). *Preventing reading difficulties in young children.* Washington, DC: National Academy Press.

Soar, R. S. & Soar, R. M. (1976). An attempt to identify measures of teacher effectiveness from four studies. *Journal of Teacher Education, 27*, 261-267.

Soar, R. S. & Soar, R. M. (1979). Emotional climate and management. In P. L. Peterson & H. J. Walberg (Eds.), *Research on Teaching: Concepts, Findings, and Implications* (pp. 97-119). Berkeley, CA: McCutchan.

Strauss, A., & Corbin, J. (1990). *Basics of qualitative research: Grounded theory procedures and techniques.* Newbury Park, CA: Sage.

Taylor B., Pearson, P. D., Clark, K., & Walpole, S. (1999). *Schools that beat the odds* (Research Report 2-008). Ann Arbor: University of Michigan, Center for the Improvement of Early Reading Achievement.

Taylor, B. M., Pearson, P. D., Clark, K., & Walpole, S. (2000). Effective schools and accomplished teachers: Lessons about primary-grade reading instruction in low-income schools. *Elementary School Journal, 101*(2), 121-165.

Taylor, B. M., Pearson, P. D., Peterson, D. S., & Rodriguez, M. C. (2005). The CIERA School Change Framework: An evidence-based approach to professional development and school reading improvement. *Reading Research Quarterly, 40*(1), 40-69.

Thorndike, R. L. (1982). *Applied psychometrics*. Boston, MA: Houghton Mifflin.

Walsh, K., Glaser, D., & Wilcox, D.D. (2006). *What education schools aren't teaching about reading – And what teachers aren't learning*. Washington, DC: National Council on Teacher Quality.

Wilson, S.M., Shulman, L.S., & Richert, A.E. (1987). '150 different ways' of knowing: Representations of knowledge in teaching. In J. Caulderhead (Ed.), *Exploring Teachers' Thinking* (p. 104-124). London: Cassell.

Appendix A – Draft Example Items from the LIKS Written Survey Subscale

Decoding:

Kyle, one of Mrs. Valcourt's first-grade students, reads the sentence, "The hot dog tasted great!" However, Greg pronounced the word *great* as *greet*. What should Mrs.Valcourt say?

A. Tell me the sound of each letter, then tell me the whole word.

B. Think, what do the first part and the last part of the word say? Now put them together.

C. Think what sound the *ea* spelling pattern makes. Now say the whole word.

D. This word doesn't follow the rules. This is the word 'great.'

Comprehension:

Mrs. Paige recently read two nonfiction books to her class. One of the books was about ants and the other about spiders. She would like her students to compare and contrast the two. Completion of which of the following would be **most** useful in allowing her students to compare and contrast the two books?

A. semantic map

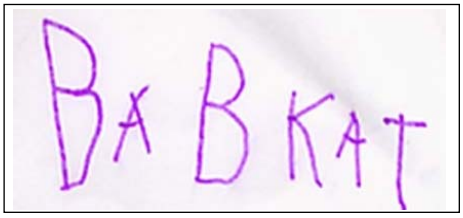
B. story map

C. KWL chart

D. Venn diagram

Writing:

Derek wrote this note to remind his teacher to bring the "book about cats."



What would experts suggest that Derek be taught about writing for the purpose of developing phonemic awareness?

A. to use lower case letters

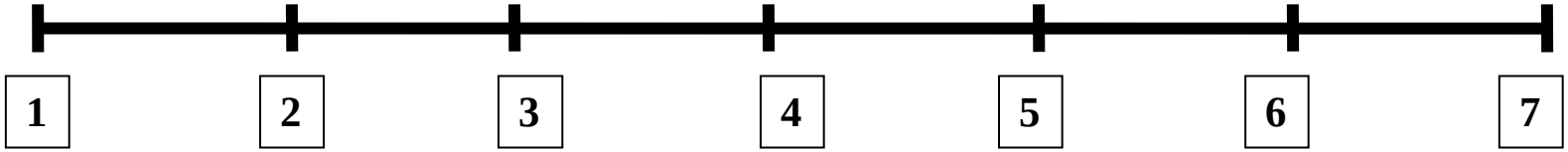
B. to listen for and write the ending sounds of words

C. how to spell the vowel digraph *ou*

D. to put spaces between words

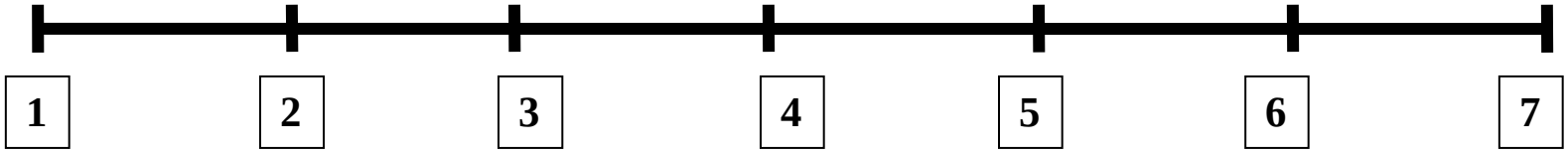
Appendix B – Draft Example Items from the LIKS Classroom Observation Subscale

Decoding and Fluency Instruction
O-DF1— Phonemic Awareness and Phonics



1	2	3	4
There is *no evidence* of effective of word work instruction. Phonemic Awareness □ *Provides no instruction in how to blend and segment speech sounds (phoneme awareness).* OR Phonics □ *Provides no instruction in how to blend and segment individual letter-sounds (CVC words-phonics).* OR Phonics □ *Provides no instruction in how to blend and segment spelling patterns and word parts (consonant blends and digraphs, vowel spellings, affixes).*	There is *minimal evidence* of effective word work instruction. Phonemic Awareness □ *Mentions how to blend and segment speech sounds (phoneme awareness).* OR Phonics □ *Mentions that students should "sound out" words. Does not demonstrate blending or segmenting processes (CVC words-phonics).* OR Phonics □ *Mentions how to blend and segment spelling patterns and word parts (consonant blends and digraphs, vowel spellings, affixes).*	There is *moderate evidence* of effective word work instruction. Phonemic Awareness □ *Explains how to blend or segment speech sounds and provides teacher guided practice of how to blend or segment speech sounds (phoneme awareness).* OR Phonics □ *Explains and provides teacher guided practice of how to blend or segment individual letter-sounds (CVC words-phonics).* OR Phonics □ *Explains and provides teacher guided practice with how to blend or segment spelling patterns and word parts (consonant blends and digraphs, vowel spellings, affixes).*	There is *strong evidence* of effective word work instruction. Phonemic Awareness □ *Provides an explicit explanation, teacher modeling, teacher guided practice, and independent practice of how to blend and segment speech sounds (phoneme awareness).* *OR* Phonics □ *Provides an explicit explanation, teacher modeling, teacher guided practice, and independent practice of how to blend and segment individual letter-sounds (CVC words-phonics).* OR Phonics □ *Provides an explicit explanation, teacher modeling, teacher guided practice, and independent practice of how to blend and segment spelling patterns and word parts (consonant blends and digraphs, vowel spellings, affixes).*

Comprehension Instruction
O-C2—During Reading



1

There is ***no evidence*** that the teacher facilitates student understanding of text during reading *(includes instruction that occurs during teacher read-aloud).*

- *Asking Questions/Discussion - Teacher asks no questions related to the comprehension of the text.*
- *Interaction—No interaction takes place among teacher and students.*
- *Engagement-Almost non of the students are engaged in the reading of the text.*
  Engagement is evidenced by:
  - On *task behavior*
  - *Eyes on text and/or teacher*
  - *Raising hands*
  - *Interaction with teacher or peers*

2

There is ***minimal evidence*** that the teacher facilitates student understanding of text during reading *(includes instruction that occurs during teacher read-aloud).*

- *Asking Question/Discussion – Teacher asks only literal questions related to the comprehension of the text.*
- *Interaction—Interaction is not focused on the text.*
- *Engagement-About 1/4 students are engaged in the reading of the text.*
  Engagement is evidenced by:
  - *On task behavior*
  - *Eyes on text and/or teacher*
  - *Raising hands*
  - *Interaction with teacher or peers.*

3

There is ***adequate evidence*** that the teacher facilitates student understanding of text during reading *(includes instruction that occurs during teacher read-aloud).*

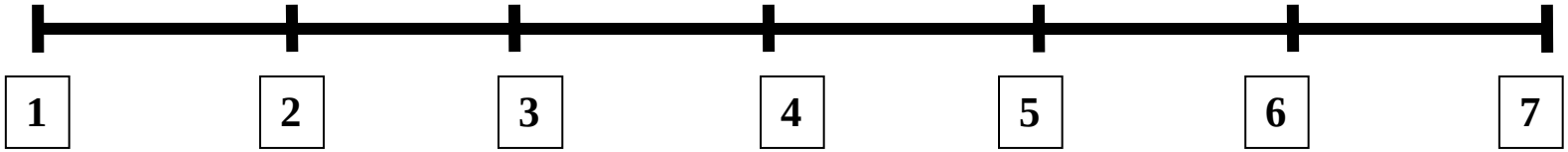
- *Asking Questions – Teacher asks a combination of literal and inferential questions related to the comprehension of the text.*
- *Interaction—Interaction is somewhat focused between teacher and students on the topic or text.*
- *Engagement—More than half of students are engaged in the reading of the text.*
  Engagement is evidenced by:
  - *On task behavior*
  - *Eyes on text and/or teacher*
  - *Raising hands*
  - *Interaction with teacher or peers*

4

There is ***strong evidence*** that the teacher facilitates student understanding of text during reading *(includes instruction that occurs during teacher read-aloud).*

- *Asking Questions – Teacher and students ask a combination of literal and inferential questions related to the comprehension of the text.*
- *Interaction—Interaction is consistently focused between teacher and students on the topic or text.*
- *Engagement- Almost all students are engaged in the reading of the text.*
  Engagement is evidenced by:
  - *On task behavior*
  - *Eyes on text and/or teacher*
  - *Raising hands*
  - *Interaction with teacher or peers*

Writing
O-W3—Writing Tasks



1

There is ***no evidence*** of students accomplishing writing tasks.

- *No writing task.*
- *No choice.*

2

There is ***minimal evidence*** of students accomplishing writing tasks.

- *Writing tasks restricted to copying, fill-in-the-blank or template driven forms.*
- *Student choice limited by strict parameters of the form provided by the teacher.*

3

There is ***moderate evidence*** of students accomplishing writing tasks.

- *Writing task involves students generating similar responses to a teacher-defined writing assignment.*
- *Teacher allows students to choose from a limited range of responses to assignment.*

4

There is ***strong evidence*** of students accomplishing writing tasks.

- *Writing task involves students generating unique responses to a teacher-defined writing assignment. (e.g. reading response, writing in a specific genre, writing as form of inquiry).*
- *Teacher allows students to choose either the topic or form of writing.*