Changes in the Observation Survey: Making It Easy to Use!

Diane E. DeFord, University of South Carolina

In the newest edition of *An Observation Survey of Early Literacy Achievement* (Clay, 2002), Clay displays a letter written to her by one of her many young fans (p. 48). In this letter, the writer suggests that she needs to write that book again (*Sand or Stones*) because she has “messed up!” While Clay will not likely rewrite the Concepts About Print task, she has been very busy writing! In the past 4 years, several new books have been published that clarify or extend aspects of Clay’s theory of emergent literacy: *By Different Paths to Common Outcomes* (Clay, 1998); *Stirring the Waters: The Influence of Marie Clay* (Gaffney & Askew, 1999); two new Concepts About Print tasks with the teacher’s guide entitled *Concepts About Print: What Have Children Learned About How We Print Language?* (Clay, 2000); *Running Records for Classroom Teachers* (Clay, 2000); *Change Over Time in Children’s Literacy Development* (Clay, 2001); and now, the second edition of *An Observation Survey of Early Literacy Achievement* (Clay, 2002).¹

When someone retires, most people ask, “Been doing some fishing?” (or traveling, etc.). But when Clay retires, the question has to be, “What did you write?” When she writes, Reading Recovery professionals take note—they want to understand what the changes she frequently introduces in her new writing mean for their teaching and assessment practices. They want to know what rationales fueled the changes (theoretical and practical rationales). This article attempts to describe the changes and rationales behind the revisions included in the second edition of *An Observation Survey of Early Literacy Achievement*.

A Brief History of the Development of the Observation Survey

Marie Clay’s research emanated from her dissatisfaction with theories of reading acquisition in the 1960s. As a child psychologist, she had taught special needs children, and she had also taught school psychologists about testing and measurement and how to work with special needs children. The major research question she began with—“How early could you see the process of learning to read moving off course?” (Clay, 1982, p. 1)—has since revolutionized our present world. Clay began her journey by observing 5- to 7-year-olds learning to read. As she tried to understand the range of differences she saw, she struggled with the question “Why do some children fail to reach their potential?” Her journey has resulted in the formation of a more complete theory of literacy acquisition—one that describes the course of literacy development and the different paths students might take.

A key outgrowth of her research was a set of highly reliable observation tasks. When she began her dissertation—

¹ In the remainder of this article, *An Observation Survey of Early Literacy Achievement* will also be referred to as the Observation Survey.
Shanahan and Newman (1997) named *The Early Detection of Reading Difficulties* (Clay, 1979/1985) as one of the 13 most influential literacy studies since 1961.

Research at the University of Minnesota, tests like the new Illinois Test of Psycholinguistic Abilities (ITPA) were commonly used. Clay disagreed with the basic assumptions underlying such tests. (Gaffney & Askew, 1999, p. x). So she revised the tools used in her dissertation and early research and published these for classroom teachers to use to enhance their observation of children’s progress. They were first published under the title *The Early Detection of Reading Difficulties: A Diagnostic Survey* (Clay, 1972). Updated versions were published in 1979 (with Reading Recovery procedures) and again in 1993 (without Reading Recovery procedures) with a new title *An Observation Survey of Early Literacy Achievement.* The most recent version of these observation tasks has additional changes other than the new title. But the basic purpose and quality of these assessment tools remains the same: observation.

The current assessment context, here and in other countries, has been to place an emphasis on the use of assessment for monitoring progress and reporting that progress. The main goals have been accountability, resource allocation, and reporting progress. There has been little emphasis on integrating assessment with learning—or to aid in instructional decision making (Hill, 1998).

The main purpose of Clay’s observation tools is to aid classroom teachers so they can integrate assessment into instruction. In her words, “educators have done a great deal of systematic testing and relatively little systematic observation of learning” (2002, p. 12). She describes the four characteristics that her systematic observation tools have with good measurement instruments. They provide

- “a standard task,
- a standard way of administering the task,
- ways of knowing when we can rely on our observations and make valid comparisons, and
- a task that is like a real world task as a guarantee that the observations will relate to what the child is likely to do in the real world (for this establishes the validity of the observation)” (2002, p. 12).

Concepts About Print, one of the tasks within the Observation Survey, grew out of the needs of the research setting. In an early research project, Clay sought to describe what a sample of 5-year-old children did as they read simple books (1982). As she observed, she carefully noted what the children were doing. She had an idea for how she might help teachers see the same kinds of responding she was noting during her research observations. She took some little books that were published in England that had similar features of layout and print but were different in content. She even took out a few pages and turned the pictures and the text upside down! She attempted to capture the early awareness children might have before they had been taught to read. This meant that the task could not ask them to read or write, and since young children cannot articulate their thinking, it could not ask them to name things or talk about more abstract concepts of print.

As the pilot task began to take shape, she used her knowledge about how to engage children in this age group as well as her knowledge about what makes a good or a bad test item. So she had children point to items and turn pages so that they could show, rather than tell, what they knew about print and how print works in books. This measure of concepts of print was used together with other tasks to help Clay describe the range of behaviors she observed in young children, such as their knowledge of letters, abilities to read continuous text, general word-reading abilities, and writing capabilities. After completing test development procedures, a total of six literacy tasks were included in the Observation Survey battery (Clay, 1972, 1993). These tasks, which first grew out of a research study, became a set of tasks that could reliably “quantify the progress of high, middle and low progress children after one year at school” (Clay, 2000, p. 18). Later this survey became the observation tool used to select children into Reading Recovery.

The survey tasks were not developed so that teachers could “test and file” the information away. Rather, they were designed “to make a teacher attend to how children work at learning in the classroom” (Clay, 2002, p. 13). They were intended to provide a learning challenge for a child that would allow teachers a unique window through which to observe chil-
Clay emphasizes a change in language toward a clearer use of terms:

- **tasks not tests**
- **strategic activities not strategies**
- **sources of information not cues**

Clay felt that if she provided a different kind of observation task, it would confront the teacher with a new kind of evidence of a child’s strengths or problems. With this perspective, and the desire to provide the best measurement conditions, we can be assured of using a sensitive observation tool that elicits reliable responses every time it is administered. In this way, she has challenged us to the very core of our theoretical beings! The evidence she forces us to see, and the way she guides us to rethink our own assumptions, has revolutionized today’s theories of how children develop, and the multiple paths they can take to literacy. For example, Shanahan and Newman (1997) named *The Early Detection of Reading Difficulties* (Clay, 1979/1985) as one of the 13 most influential literacy studies since 1961.

### Overview of the Changes in the Observation Survey

Several obvious changes may be seen in the second edition of *An Observation Survey of Early Literacy Achievement*. It is almost twice as long. While part of this length is due to a slightly larger font, there are many format changes, more examples, and other additions that provide clarifications, stronger rationales, and necessary updates to keep abreast of new information from current research and literacy practices that have emerged in the last decade. The most significant change is in the new norms that have been added and the explanatory information provided to support classroom teachers in using the new normative data.

One of the greatest threats to reliability (consistency and accuracy of a task’s performance) is related to how people administer an instrument. Consistency is increased if the explanations and descriptions that accompany each task are clear and complete to weed out faulty assessment practices. So one type of change that is evident in the new edition of the survey is in the clarity of the task descriptions and explanations. I will address each of these changes and provide explanations about how these changes impact the use of this survey for assessing early literacy achievement.

### Running Records

Clay has clarified or provided additional rationales, for example, on the use of running records. In Chapter 5, Clay more clearly explains why teachers need to practice taking running records on average children (p. 54) to become more skilled at recording, scoring, and interpretation before working with higher- or lower-progress readers. She addresses why we should avoid using a preprinted page of text (pp. 53–54) and adds how to mark a verbal appeal for help from the child that is turned back to the child for further effort. She directs us to say, “You try it” (recorded as Y). The larger font and the examples provided make it easier to see and understand the standard marking procedures. She also has changed the recommended time to wait before a Told (no more than 3 seconds). On page 59, Items 6, 7, and 8, she discusses how to capture a Told (T) and Try That Again (TTA) and adds how to mark a verbal appeal for help from the child that is turned back to the child for further effort. She directs us to say, “You try it” (recorded as Y). The larger font and the examples provided make it easier to see and understand the standard marking procedures. She also has changed the recommended time to wait before a Told (no more than 3 seconds). On page 59 in the second edition, Clay says, “If the child baulks, unable to proceed,” wait “no more than about” 3 seconds. However, if the child is problem-solving—and it takes a noticing teacher to observe what the eyes are doing as well as what the mouth is saying—then you may want to allow a little more time. I found the samples of running records and the guidance for marking and scoring to be consistent and much improved.
Clay has provided one example on page 63 that attempts to clarify the scoring of errors and self-corrections. In Item 2, “There is no penalty for trials which are eventually correct,” she provides three examples. Example C is the item to note, and below this example, a further clarification is given (this also relates back to Item 3 on page 58). Items A and B are the same as in the previous edition:

When the child substitutes one word for another word, or if there are multiple full word attempts on a single word, and the child subsequently corrects, the scoring directions say it counts as one self-correction, but no error. Example C, however, is new.

**Example C**

<table>
<thead>
<tr>
<th>Child:</th>
<th>f-</th>
<th>fet</th>
<th>✓</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text:</td>
<td>fright</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Score:</td>
<td>—</td>
<td>—</td>
<td>✓</td>
</tr>
</tbody>
</table>

Example C deals with attempts to solve the word when the attempt is a nonsense word but the final word is correct. In this instance, there is no initial error, so there is no self-correction (a ✓ is used within the record to show the correct word was said).

Later on page 73, Clay discusses signs of progress in older readers and provides an example of an interrupted response where the child begins to say one word which might reasonably follow in the sentence but interrupts that response and produces the correct word. She writes: “A similar kind of thing happens when the older reader corrects what might have been an error [italics added] before giving the whole word as in m…/parents; gar…/ground; d…/tied.” Thus, she makes the distinction between when the child is solving a word which is not a self-correction and when the child is interrupting the pronunciation of an anticipated word and making a self-correction. Teachers who may have been counting all partial attempts that look visually discrepant as self-corrections now need to distinguish word solving or attempts to sound out from false anticipations. All of these explanations in how to score errors and self-corrections bring about greater consistency (reliability).

On page 69, Clay provides three questions to guide our interpretations of the running record to help us think about what influenced the reader. Of note is the point she makes about structure (syntax). The question is “Did the structure (syntax) of the sentence up to the error influence the response?” She then argues that “if the error occurs on the first word of the sentence it is marked as positive for structure if the new sentence could have started that way.” This is a change in how we handled structure in the past, when we did not circle structure if the new sentence could not have been a continuation.

Another subtle but important change is illustrated on the running record form example on page 71. At the bottom of this example, a comment states, “Read slowly with some intonation.” On page 61, Clay explains that at the end of the record, we should now comment on our overall impression of the reading. For example, we might ask “Was the reading done at a good pace, or was it slow, or too fast? Are things in balance or out of balance in your judgment?” This comment about how the reading sounds is also included in the Observation Survey Summary Sheet (pp. 124–125). These changes in the forms help us to better reflect on the complexity of the reading process with a focus on fluent reading and useful strategic activities on text, words, and letters.

Clay has specifically addressed assessment of comprehension with some comments and cautions about adding retellings or comprehension questions to a running record. “The reliability and validity of these assessments are not improved when teachers cannot agree on the scoring or when what teachers do is non-standard, like asking questions which differ in content, form or purpose” (p. 61). Clay also provides guidance about interpretation of running records (p. 70) with questions teachers can ask to explore patterns of responding. She also addresses common faults she has observed in the analysis of running records. You may need to take some time to explore changes in figuring accuracy and self-correction ratios (p. 66). Clay’s goal is to improve reliability and validity in the use of running records through changes aimed at greater standardization in administration, scoring, and analysis. She has also helped classroom teachers better monitor children’s progress and fashion appropriate instruction (see pages 72–81).

**Concepts About Print**

Chapter 4 isolates Concepts About Print from the other tasks as a chapter by itself (see pages 37–48). The two new color versions of the task booklets, No Shoes and Follow Me, Moon, are included with the original Sand and Stones alternatives in this edition. The two new children’s books were also used in the normative research.
study completed by Reading Recovery trainers at the National Reading Recovery Centre in Auckland. Achia’s critique (from the child’s perspective!) of the Concepts About Print task is a must read for everyone (see Clay, 2002, p. 48)! 

Other Tasks
Some changes in each task area may need further study at professional development sessions (see Chapter 6, pp. 82–120). In the Letter Identification section, Clay integrates more research in this area (pp. 82–90), relates student scores to norms taken at four points in time (p. 87), presents research on easy-to-learn letters and hard-to-learn letters (p. 89), and offers some cautions on assessing letter-sound relationships that will prove very helpful to us (pp. 88–90).

In terms of the Word Reading task (pp. 91–97), Clay differentiates the one she has consistently used in New Zealand (Clay Word Reading Task) and those that are grouped into other first-word reading tests (Canberra, Ohio, Duncan). She describes the Clay and the Duncan as equally helpful and as optional alternatives in statistical tests. For our purposes, however, we will still use the Ohio Word Test in Reading Recovery in the United States. Again, Clay has provided norms at four points in time and compares the Clay Word Reading Task with other common word and reading tests used in the different countries (Burt, Neale, STAR, and Duncan) available for classroom use.

The Writing Vocabulary (pp. 97–111) section describes how to collect and

Follow Me, Moon, one of the two new versions of the task booklet in Concepts About Print, allows children to identify what’s wrong with this page of upside-down type. The letter below shows that one young reader noticed.

Dear Marie,
You messed up the book. Write it again.
Love, Achia

---

2 The Ministry of Education drew a nationally representative sample of schools, and children were randomly selected in each school. Dr. Ken Rowe of the Australian Council for Educational Research analyzed the data and acted as a consultant on how it might best be presented and used.
rate three classroom writing samples and how to administer the Writing Vocabulary task. It includes helpful samples to illustrate the range of writing one might get from these two tasks. In terms of the Writing Vocabulary task, she adds two cautions (p. 104). First, she cautions against prompting for words in a series (can, man, etc.). Second, she argues against dictating from a list of words which children might have met in reading books or vocabulary lists used for teaching. She still emphasizes prompting words in pairs, like go or me for this task. In the revised edition, the word I is scored as correct even if it is written in its lower-case form (p. 106).

The administration and scoring of Hearing and Recording Sounds in Words has a few notable changes (see pp. 111–120). She places this section in today’s context by adding a subheading “going from phonemic awareness to letter-sound relationships” (p. 111). She provides the backdrop of the literature of the 1970s, out of which this task was derived, and then gives information about this task, phonemic awareness, and orthography (pp. 111–112). In terms of the sentences themselves, Form B has a provision for Mum as well as Mom. She also provides examples for each sentence that have some special coding problems (p. 117). The biggest changes in this section are in the scoring examples provided on page 116 for substitutions. Clay has resisted giving many examples of substituted sounds in previous revisions. The examples that brought about the most heated discussion in one professional development session was the substitution of caming for coming and bas for bus, in which the vowel substitutions count as correct. Clay’s rationale is that the indefinite article a is pronounced like the u in bus and so this might be an appropriate substitution for emergent writers. The spelling research on older students documents how the vowel substitutions will vary widely as students come to grips with vowel patterns in multisyllabic words and the pronunciation of the schwa sound in English. It helps to remember we are trying to stand in the young child’s shoes, looking at print from the novice’s perspective!

New Norms
Perhaps the most significant change in the second edition of An Observation Survey of Early Literacy Achievement is in the research that has been done most recently in New Zealand to produce new norms at four different points in time (see Appendix 1, p. 148–159). Also included are new ways of reporting information about the sample (such as box plots and graphs). The new norms were developed from a random sample of New Zealand children (n=796) who were between 5 and 7 years of age. Four children were randomly selected from each age group (5.0–5.5, 5.1–6.0, 6.01–6.5, 6.51–7.0) in 199 representative schools. The results of this large-scale testing indicate how the educational system in New Zealand has changed since the original norms were created in 1968 and 1978, given the shift in scores in a positive direction across tasks. Having these norms broken out by age group also helps teachers find the best sample of students for individual or group comparisons in a “temporary progress category” (p. 121). The use of these norms and stanines are woven throughout each assessment task in a helpful way. Clay gives a strong rationale for the use of these stanines and ways to interpret the information gained from their use.

Putting the Observation Survey and Its Contributions Into Perspective
The new edition of An Observation Survey of Early Literacy Achievement will help reading specialists, Reading Recovery personnel, and classroom teachers understand the importance of observing and noticing a child’s processing in reading and writing.
Throughout the second edition, Clay has sought to place more emphasis on how children are processing within and across reading and writing. In Chapter 2, she goes into greater depth about both the reading and writing processes, and how children process information in print by using many sources of information. This particular chapter increased from two to nine pages. In addition to new examples, the text incorporates new research and new thinking that is drawn from Change Over Time in Children’s Literacy Development (Clay, 2001) and By Different Paths to Common Outcomes (Clay, 1998). These additions allow Clay to make a stronger case for studying patterns of responding and noticing gradual change over time. A new section on page 21 gives information about why it is important to have a blank page for writing, and there is also a section in which Clay discusses seeing print from two vantage points (as a reader and as a writer). These changes offer a better balance between reading and writing process discussions in the second edition.

As I have studied this new volume, I have admired how change is not left to chance. Just as Reading Recovery: A Guidebook for Teachers in Training is a constant resource for teachers as they do their daily teaching, An Observation Survey of Early Literacy Achievement is a constant guide to classroom teachers and Reading Recovery teachers alike. It keeps teachers focused on observation and helps them in monitoring children’s progress over time to capture the natural fits, stops, and surges of the emergent learner. The new elements and examples provided in the 2002 edition also remind us to limit conclusions to those supported by evidence as a way to maintain objectivity and avoid bias—the biases of our own preconceived notions about how children should perform. Changes in this document are the result of a careful examination of current research across a wide range of disciplines and perspectives as well as research specifically designed to further document the reliability and validity of each task in the Observation Survey. As education professionals, we should carefully examine and act upon these changes and adhere to standard procedures of administration and guidelines for interpreting student responses on these sensitive observation tasks.

Marie Clay’s contributions to early literacy have shown her to be, as Richard C. Anderson wrote, “one of the most remarkable educators and scholars of the twentieth century” (1999, p. vii). This new edition of An Observation Survey of Early Literacy Achievement will guide our learning about children and their progress for years to come. It represents a sound contribution to the field of testing and educational measurement, and the revisions to date effectively demonstrate its reliability and validity.

In the face of recent criticisms and a shift to high-stakes assessment, it is important to know where the Observation Survey stands in comparison to other assessments commonly used today. The strength of the Observation Survey as an assessment tool gives readers a way to respond to recent attacks of its scientific basis and appropriateness. (For a thorough treatment of this topic, see What Evidence Says About Reading Recovery, [RRCNA, 2002]).

The only data that count in the determination of a scientific basis for tools of assessment are statistical tests of reliability and validity. The strength of the tool is determined through the processes of testing the degree of consistency and accuracy with which a test performs (reliability) and the degree to which it measures what it says it measures (validity).

Establishment of standardized procedures and conducting carefully designed research to provide norms of performance also feed into the effectiveness of an assessment tool. The reliability, validity, and standardization have to be checked against the purpose the test was developed to serve. For An Observation Survey of Early Literacy Achievement, the purpose was to describe ranges of difference in children’s reading and writing performances, to seek evidence about why children fail to realize learning potential, and to understand the course of literacy development and the different paths students might take. Given these purposes and Clay’s adherence to the above formal tests of reliability and validity, An Observation Survey of Early Literacy Achievement has to be judged as having a strong scientific base and to be an appropriate assessment tool to describe and monitor children’s early literacy achievement.

In today’s high-stakes assessment context, understanding some of the deci-
"I value highly an approach to research which attempts to capture the ways in which children change over time in a reading program, an approach too rarely taken in reading research."

Marie Clay, 1982, p. xiv

The decisions Clay had to make when developing these tasks is important. (For a complete review of the reliability and validity research completed on these tasks, see Clay, 2002, pp. 159–165.) A quick review of some of these decisions should help define the larger context surrounding the Observation Survey. The following comments are drawn from Clay’s scholarly writing to provide information to counter criticisms that might be leveled in today’s school context.

• The vast majority of experimental studies follow paradigms of group comparisons and manipulation of controlled variables; the limitation of studies based on group comparisons is that individual data are aggregated and reported as group means. Clay states that “research studies hide the complexity of child learning by reporting group averages, but a pooled average may not describe any individual” (1998, p. 226). When the intent is to describe individual differences, then a different research design must be employed. But this design can be held to very high research standards without compromising the findings or losing sight of the original purpose.

• Closed tasks which are typical of many tests provide data that represent final outcomes of children’s learning (achievement). These “error-free end results” hide false starts, half-correct processes, and self-correction (1998, p. 257). Clay’s research design includes both open and closed tasks. This allows for observation of these emerging, tentative behaviors to detect the variability of individual paths to literacy achievement. Even though some of the Observation Survey tasks result in what is called a ceiling effect, it is better to think of these tasks as measuring knowledge of a closed set of items (a certain number of letters or phonemes). When students respond correctly, we have data that they have reached an optimal level of performance. Several tasks qualify as criterion measures to assess children’s developing knowledge in limited domains (Letter Identification, Concepts About Print, Hearing and Recording Sounds in Words). These tasks are useful for informing instruction of 5- to 7-year-olds.

• The Observation Survey has high construct and face validity and high reliability determined through statistical tests. The measurement error within any task is greatly reduced with individual administration and when accompanied by standardized procedures. These tasks yield reliable rankings but ceiling effects limit usefulness beyond the early reading stage. The new edition provides normative data on a more extensive sample of students between 5 and 7 years of age with New Zealand norms. These will be followed soon by new U.S. norms. The need to update the validity and reliability of the tasks for
the United States is in response to Reading Recovery scholars and critics.

- Studies discussed in Observing Young Readers (Clay, 1982) extend or verify elements of theory that followed rigorous standards. The Observation Survey reports results of large-scale tests of literacy tasks that were an outgrowth of previous investigations and theoretical work. “I value highly an approach to research which attempts to capture the ways in which children change over time in a reading program, an approach too rarely taken in reading research. It is not enough to administer achievement tests before and after a program or to probe deeply into the reading process. An important additional source of research information lies in accounts of what changes occur in what children in what sequences as they learn in their classroom programs. As a researcher this has interested me; it is, of course, what concerns the teacher” (Clay, 1982, p. xiv).

- Clay did not intend to develop norm-referenced group achievement tests. The psychometric principles used for constructing clinical assessments of individuals, not groups, were applied to the tasks in her test battery. Group tests are not useful if the baseline measures for assessment of longitudinal development are taken during the emergent phase of any new achievement. As stated by Clay, “Powerful statistical analyses have shown that these procedures which permit more detailed recording of individual responses than a normative test, nevertheless have proved to be sound measurement devices” (1982, p. 6).

- It is easy to arrive at false assumptions from averaged data based upon superficial or highly selected observations, and assumptions may be the result of oversimplified, logical analyses of tasks that bear little relationship to the ways individual children learn. Standardized tests make all of these errors.

Conclusion
As I have used the Observation Survey over the past several decades (I won’t say how many!), I am struck by the strength of these assessment tasks on several critical fronts. In a review of test instruments for assessment, Croft, Strafford, and Mapa (2001) suggested that tests should be evaluated based upon four criteria. I will close with a brief discussion of An Observation Survey of Early Literacy Achievement in light of each of these criteria:

- the theoretical soundness of the total survey,
- the high content validity of the survey tasks as measures of early literacy achievement,
- the ease of use, and
- the high quality of each task for use by classroom teachers, reading professionals, and Reading Recovery personnel.

The Observation Survey tasks were developed with care to ensure that each task would be a robust measure with items that are highly relevant for use with the emergent reader and writer.

Theoretical Foundation
The tasks themselves are backed by existing and recent publications that describe the theoretical underpinnings of these tools and their strong base in research across four decades (Clay’s By Different Paths to Common Outcomes [1998], Change Over Time in Children’s Literacy Development [2001], and Becoming Literate: The Construction of Inner Control [1991]). Some tests currently in use today lack a conceptual or theoretical model, or the model on which they were built is outmoded.

Content Validity
Few measures publish adequate information about the origins of assessment tools and their development. It may be that test developers expect that users will be able to determine the relevance of the tasks or items to their curriculum needs. Clay’s Observation Survey has careful descriptions of each task and how the information relates to development of literacy and instructional implications. Teachers need to be selective when choosing any tests to make sure they provide sufficient information to determine content validity. It may be that the particular approach to curriculum on which a test is based has changed.
Ease of Use
After looking at the quality of the tasks, it is important to see if there is enough guidance for teachers in how to interpret the results of the assessment. If the test only provides tables with scores, but little discussion of individual scores or what score patterns mean, it will make it difficult to use the tool. If the assessment tool can only be used by trained professionals, it is less useful. The second edition (although it is the fifth version in a historical sense; see Clay 2003, p. 164) of An Observation Survey of Early Literacy Achievement has incorporated many helpful samples, careful descriptions, and very informative discussions that make it a highly utilitarian text. Teachers not only get valuable information about each student and a classroom monitoring scheme, they also increase their professional knowledge about the reading and writing processes and the acquisition of reading and writing in young children.

Robust Quality of Tasks
In developing a test, each target area of assessment has to be weighed as to how many items to include. Should it contain a large sampling of items and run the risk that few teachers will be able to use the longer task? Or should the test developer select only the most relevant items, knowing that the sampling is more limited? Good diagnostic tests take a lot of time and a great deal of effort (to say nothing of money!) to develop and publish. They need to test each task to find the best balance. But it may be that some diagnostic tests are deliberately limited in what they attempt to assess or they may measure it at a basic level only—and this may not suit a teacher’s context. In some instances the skill that is being measured may not have changed, but our understanding, beliefs, and assessment practices may have. The Observation Survey tasks were developed with care to ensure that each task would be a robust measure with items that are highly relevant for use with the emergent reader and writer. This survey was also developed so each task would be easy to use in the classroom setting.

Few of the assessment tools used today can offer such a rich theoretical foundation or description of each task’s origins and use as An Observation Survey of Early Literacy Achievement. These tasks may appear simple, but they represent a significant innovation in educational practice for instruction and research, as well as for program evaluation.

References


