# WILEY

# INTERNATIONAL
# Reading Association

# Skills management systems: a critique

DALE D. JOHNSON
P. DAVID PEARSON

Dale Johnson is involved in teaching and research related to beginning reading at the University of Wisconsin, Madison. David Pearson is also concerned with teaching and research in reading at the University of Minnesota, Minneapolis.

THE NOTION of accountability in education seems to be gaining more and more popularity. As more states pass accountability laws and conduct statewide assessments of reading achievement, and as the federal funding programs begin to demand achievement growth as a contingency for refunding, one cannot help but conclude that accountability is here to stay.

The question that naturally arises is, "What is it that teachers of schools ought to be held accountable for?" Because of our educational system's penchant for testing and assessment, it is likely that some form of assessment will provide the data base for future accountability schemes. Given this constraint and the focus of this paper, the question of accountability in reading becomes, "Does a skills management system provide an accountability scheme that is preferable to some alternative scheme, for example, a standardized reading test or an index of the number of volumes checked out of school or community libraries?"

The answer to that question can best be achieved by analyzing the nature and underlying assumptions of a skills management system for reading instruction. That analysis can then be compared with one's own conception of the nature and goals of reading instruction in order to evaluate the utility of a skills management system (SMS) as an instrument of accountability.

The commercially available systems and the locally developed systems with which we are familiar share these components: 1) a sequentially ordered set of behavioral objectives for the various reading skills monitored by the system, 2) a set of subtests (or a set of test items) with

**757**

one or more items designed to measure each objective, 3) a rule or set of rules for deciding what level of achievement constitutes mastery of each objective, 4) a resource file listing specific workbook pages, Ditto masters, games or kits, and (hopefully) teaching strategies which teachers can use to provide instruction and practice for children who have failed to attain mastery of specific objectives, and 5) a method of reporting to teachers which students have or have not mastered which skills.

## Purported advantages

With these resources at hand, teachers are supposed to be able to individualize instruction to a greater degree. That is, children who are making excellent progress in skill development (in other words, pass the tests) will not have to endure the drudgery of instruction in matters they have mastered. They can spend their time at more productive tasks such as independent reading or special research and reporting projects, while the teacher is secure in the knowledge that they have learned the basics. On the other hand, those who demonstrate less than optimal progress can participate in exactly those activities which they need in order to bring their skill profile into line. In short, wasted instructional effort, because a task was too easy or too difficult for a particular child, is eliminated. Furthermore, multiple opportunities exist for cross-age or cross-grade grouping for short-term instruction in specific skills.

The system also provides teachers with a meaningful way to discuss student progress with parents or with children themselves. No longer will teachers have to be embarrassed by telling parents that a child "has trouble with reading." Progress can be reported in a more specific, meaningful way. (We know of one school where the progress made on specific skills is checked off on the report cards sent home to parents.) Likewise, in conferencing with children, a teacher can tell them that they will be in group X for N number of days. There is a little light at the end of the tunnel.

There are two other alleged advantages: 1) an SMS allows everyone—teachers, administrators, parents, students—to operate within a clearly established set of guidelines concerning what is to be done, how it is to be done and how you know whether or not you've done it, and 2) global, diagnostically useless evaluations are replaced by specific assessments which clearly prescribe instruction for a particular child or group.

There are at least six things that bother us about skill monitoring systems: 1) their psycholinguistic naiveté, 2) their "assembly-line" underpinnings, 3) their concern for skill at the expense of interest, 4) their advocacy of sequencing separable reading skills, 5) the validity of their assessment instruments, and 6) the very notion of mastery itself.

## Psycholinguistics and reading

The first point we wish to consider is the psycholinguistic naiveté evident in skills management systems in reading. We know that language systems—the phonology, grammar and lexicon—are interdependent. In essence, language is indivisible; yet SMS's would seem to fractionate it and destroy its essential nature. Because of the interdependence of the language systems, there is really no possible sequencing of skills—a point we will take up later. Psycholinguists and others have argued that the division of the reading process into component skills is unrealistic because the reader rarely uses all the information available in the text. Printed materials contain a great deal of re-

dundancy: morphemic, syntactic and graphic.

If you consider a sentence like "The girls were walking with their friends," there are three indicators of plurality which children might use to understand the passage. Yet, in some skills monitoring systems, a certain stress would be placed on testing and teaching the final *s* plural marker.

In terms of graphemic arrangements, we know that certain letters or clusters only occur in certain positions or only are found before or after certain other letters. Thus to teach or test correspondences in isolation—as some systems would seem to suggest—ignores the linguistic constraints that can enhance one's reading effectiveness. If we confine our attempts to understand reading solely by attending to observable phenomena, we then view reading as a direct response to graphic stimuli. This ignores the existence of underlying language competence. It ignores the fact that the deep structure a reader attaches to a passage may be widely variant from the writer's deep structure.

These divergences may be caused by any number of factors, including different experiential or conceptual background, particular word connotations, or the understanding of idiomatic expressions. One of the authors taught English as a second language in Nigeria for several years. (We are sure that any of you who have taught English as a foreign language quickly learned the confusion that arises, for example, over idioms.) He remembers a 14 year old boy named Musa who entered the teacher training college with a limited knowledge of English, but with a considerable amount of playfulness. Musa formed a very perplexed look when told to "get on the ball."

This fragmentation of reading inherent in all skills monitoring systems denies the nature of language.

In some of the systems, phonics is regularly taught apart from the reading lesson, whatever that may be. Yet, we know very well that some children can read well but do poorly on phonics exercises, while others do the reverse. Reading materials that would seem to make more psycholinguistic sense than those under consideration here would be materials which involve natural language, predictability of language, and meaning which relates to the learner's experience.

## The engineering syndrome

Skills monitoring systems may be viewed as the "inappropriate applications" of the engineering rationality. We think the factors involved in learning to read are too complex to be dealt with through assembly-line thinking. Arthur Brown has reflected on the human quality and accountability (1973):

> On the matter of taste, I must confess to a certain antipathy to the mechanistic orientation of the movement. People seem to disappear or move to the wings of the educational stage, while tape recorders, file cabinets, computers, records, models, and flow charts move to the center. The terminology leaves me cold: inputs, outputs, feedback, systems analysis, delivery of educational services. Finally, the philosophy: pure scientific realism. If the proponents of the movement do have a theory of man and a theory of reality that are supportive of their educational theory—and often this is not the case—it is that people are merely machines, only more complex; that they are nothing more than products of their conditioning; and that all things, including human qualities, are objectifiable, quantifiable and predictable.

Proponents of skills monitoring systems proclaim their concern for individualization but it seems to us that this is a superficial humanism. These systems tend to be conservative—even reactionary—and designed to meet the old goals in pseudo-sophisticated new ways. In our opinion these systems stress con-

tent, not process, in the framework of a tightly organized structure, leaving little room for incidental learning. They seem to be in effect a system of surveillance that potentially could restrict spontaneity, creativity and innovation. Furthermore, they tend to place the blame on children. Children fail, not teachers or systems. We question the applicability of the engineering syndrome with its assembly line techniques to the education of young children.

**Skills versus interest**

A third point which seems to merit consideration is the apparent over-concern of the systems with reading skill rather than reading interest. The SMSs confound the question of one's ability to perform on tests with one's willingness to read. Recently some would say that a concern for lifelong reading is unwarranted because of our advances with media. We think it is imperative that we be concerned with lifelong reading. Statisticians tell us that many people learn to read in school but rarely read again once they graduate. Besides the immense amount of pleasure and knowledge to be gained in reading, basic literacy is indeed a prerequisite of survival. You may remember Danny Kaye's song from *Hans Christian Andersen*:

> Inch worm, inch worm
> Measuring the marigolds
> Seems to me
> You'd stop to see
> How beautiful they are.

The point is that with skills monitoring systems we can become so concerned with the observable and measurable that we lose sight of things *not* so readily observable. We become concerned with the trivial. We think our principal goals should be preparing children who *want* to read and *do* read. We are afraid an overemphasis on "skilling and drilling" will interfere with this goal.

Skills management systems are not the first mechanisms to use a scope and sequence of reading skills for purposes of deciding what skills ought to be taught when. Basal readers, implicitly or explicitly, have used a scope and sequence concept for decades. SMSs, however, may be the first to assess skill mastery in such a demanding and systematic fashion.

The whole notion of a sequence or hierarchy of skills is, at best, a pedagogical convenience. While the idea may appeal to our sense of logic (just as we think of driving a car or riding a bicycle as a complex of sequenced subskills), there is precious little evidence to support the existence of separate skills, let alone separate skills which can be placed into a sequence of hierarchy. For example, the intercorrelations among the decoding subtests of the Stanford Diagnostic Reading Tests are consistently between .50 and .70 (Karlsen and others, 1966). In the area of comprehension, as it is measured by tests, factor analytic studies tend to find that ten to twenty subtests typically yield three to five clusters (Davis, 1941; Spearritt, 1972).

The separate skills notion may be a function of student achievement levels. For example, Guthrie (1973) found much lower intercorrelations among subtests for low achievers than for high achievers. Aulls and Pearson (1974) found that an a priori logical ordering of expected performance on decoding skill subtests was much more accurately reflected in the actual rankings of low and average achievers than in the rankings of high achievers.

Examine any three or four skills monitoring systems and you will realize that there is no agreement as to sequence. Some begin with the alphabet, proceed through the consonants' sounds alphabetically and eventually teach some vowels, while others begin with rhyming elements

and graphic shapes and dwell on the colors of the rainbow before looking at a few words and some letters. When sequences are committed to paper, the message teachers get is that there is some logical rationale for such sequences. Is there any data to show that one moves along a smooth path of learning while absorbing measurable amounts of knowledge over specified time periods?

We believe that a SMS has an obligation to its users to demonstrate that its scope and sequence of skills is more than a pedagogical convenience; that is, empirical tests should be run to verify, amend or revise the sequence that is offered.

One avenue of investigation that seems plausible is to investigate the possibility of mini-sequences. For example, it might' be possible to demonstrate that three or four subskills need to be ordered sequentially for instructional purposes (for example, auditory concept of a syllable, visual concept, dividing rules, pronunciation rules) but that that particular mini-sequence is relatively unrelated to another set of subskills.

### Validity of instruments

SMSs assume that the tests provided for measuring attainment of objectives are valid indices of the skills at issue. Yet we are unaware of any documentation which would suggest that the subtests in any SMS have been validated by relating subtest performance to any generally accepted measure of a real reading task. We will grant, for the moment, that it is not feasible to test all possible subskills with one child at a time. However, if a SMS uses paper-and-pencil measures of a non-paper-and-pencil task, then it ought, at the very least, to validate those measures by administering group and individual tests to a small sample of students. We are not convinced that

identifying the initial consonant "f" from the distractor set *f, t, v, r* when given an oral stimulus is the same as saying /f/ when seeing "f," or reading sentences which contain words starting with "f." However, we could be convinced by an experiment which attempted such a validation.

Do the items on some SMS tests actually measure what they say they measure? In one SMS test purporting to measure phonic ability (the ability to attach sound to spellings), the directions read as follows: "Given real or nonsense words pronounced by the teacher the child is able to give the initial (or final or medial) letter." Such tests are measuring spelling ability (encoding), rather than decoding—the ability to apply letter-sound correspondences. In a recent investigation correlating children's encoding and decoding responses to synthetic words, Stevenson and Johnson (1974) found that children's oral and written responses to the same word differed nearly 50 percent of the time. For example, a child may pronounce *bame* /bem/ and spell /bem/ baym.

In one SMS, comprehension is measured by requiring the child to read silently along a passage *while* a teacher reads it orally. How do we ever know what is being measured with such a task?

There are other minor validity issues, most often peculiar to specific SMSs. For example, test labeling appears to be a problem in one SMS where the root word test really measures knowledge of prefixes, suffixes and inflections (the distractors all have the root in common, only the affixes vary across distractors).

However, the other major validity assumption is that mastering all these separate skills has something to do with reading. In short, what is at issue is whether or not mastering a specific skill improves a child's ability to read or comprehend running

text. Notice that this is slightly different from the previous issue which asked whether or not assessing the skill as a group administered paper-and-pencil task was a valid substitute for assessing that same skill as an individually administered reading task. The present question is whether or not mastery of that skill, regardless of how it is assessed, transfers to something most of us would like to call "reading."

The question is, in principle, empirically solvable. Practically speaking, one probably could not determine the contribution of each individual skill. However, one could determine whether or not mastery of a set of subskills contributed to increased oral reading fluency or comprehension of written discourse. Such a validation study seems critical to the whole notion of SMSs. Why bother with the whole thing if there is no payoff in the criterion task (reading) which the whole system ought to be trying to improve?

**The mastery notion**

Most SMSs suggest that a score of 80 to 90 percent correct on all the items in a subtest is adequate in order to assume mastery on the part of an individual child. Such levels of performance appear to stem from work in the area of programmed instruction and work with college students in education psychology courses (Bloom, 1968). As with the other assumptions of SMSs, there appears to be little empirical foundation for such levels.

We must question the confidence which can be placed in particular skill measurement. If a child does well on three or four items, or if he does poorly on the same three or four items, which purport to measure the secondary sound of the *au* vowel cluster, for example, can we feel any assurance of such mastery? If a child

scores at the 80 percent criterion level, what does it really mean?

Criterion-referenced test advocates often argue, in pointing out the usefulness of their measures, that survey tests of reading achievement suffer from the fact that so many types of skills are assessed in such a test that one cannot infer that two students who receive the same raw score really performed equally well. Criterion-referenced tests which set a priori standards for mastery fare only slightly better than norm-referenced survey tests. For example, Terwilliger (1972) points out that on a ten item test with an 80 percent mastery level, there are 56 different combinations of correct items that will yield mastery. On a twenty item test, there are 6,196 such combinations, and on a fifty item test, 13,432,735,556.

A more critical problem is the issue of what it means to master a skill. As stated earlier, the whole notion of test validity is critical here. Are the test items really representative of the skill as it is to be used? Also, while the concept of mastering a word attack skill makes sense to us intuitively, the corollary concept of mastering a comprehension skill (whatever that is) makes no sense whatever. We can visualize a situation in which a teacher might decide to cut Johnny or Susie off the final consonants skill activities (although one must be careful about an allegedly mastered skill), but we can't imagine why a teacher would stop all main idea or multiple meanings activities simply because a child answered eight of ten such items correctly. If you show us a child who has mastered the level X main idea test, we can demonstrate his or her lack of mastery simply by increasing the conceptual difficulty of the words or contextual relationships. Comprehension is, by its nature, an ongoing, never-ending process. It can have no

precise starting or stopping point. It is pervasive to all reading and to all verbal discourse.

In summary, can SMSs provide a reasonable framework for establishing an accountability system for teaching reading? The answer at this point is *no*. There are too many unverified assumptions underlying SMSs. In principle, SMSs may be the ideal alternative for assessing student progress in reading. But, until some basic research and evaluation of the systems is conducted, they must be viewed only as one of many alternatives (and a very expensive and time-consuming one, at best.)

At a more personal level, as parents and taxpayers, we wouldn't want to be put into the position of accepting the conclusion that a teacher who had nurtured our children to mastery on each and every skill in the system had taught them to read. As teachers, we wouldn't want to be put in the position of being criticized for having children we know could and did read but who achieved below mastery on several skills. Likewise we would feel uneasy about those who had mastered all the paper-and-pencil tasks but could not and did not read.

We are in the embarrassing situation, typical of persons who criticize something, of having more to say about what we think should *not* be done than we have to offer about what should be done. We can state, however, what we think is preferable to a SMS accountability scheme, given present alternatives.

At this point we would prefer a standardized reading test which assesses reading achievement in the most global sense to an assessment of specific subskills. Global measures have, at the very least, the virtue of requiring that the child *read* in order to demonstrate achievement. Furthermore, they require that the child integrate a number of allegedly separate skills in order to perform the tasks required.

This preference stands even though we are very skeptical about any standardized measure of achievement. We have been involved in a great deal of first-hand frustration in failing to demonstrate that a program (which everybody—students, teachers, administrators, and parents—*knows* is successful) can produce significant growth in student achievement. Perhaps it is a function of strong home environmental factors. Perhaps it is a function of the norming and data reporting procedures of the standardized tests. We're not certain.

While it suffers from problems of reliability and objectivity, another alternative might be some combination of informal oral and silent reading inventories. Before they could be made useful, however, their methods of assessing comprehension would have to be modified. Again, such measures come closer to what we think reading is all about than do minute assessments of several skills.

Another area, though it is messy and somewhat subjective, that may ultimately provide an index of reading development concerns attitudinal factors. Measures such as amount of free time spent reading or the number of books checked out of the library may provide countable indices of effectiveness.

Accountability appears to be a part of the future of schools and teachers. There are some alternatives available for evaluating teacher effectiveness in teaching reading. Our preference lies with those alternatives that examine the reading act in its global rather than its atomistic facet. A great deal of hard-headed research and development needs to be done before we could ever be convinced to change our preferences. Educators at all levels ought to be actively involved in conducting that research.

## References

Aulls, M. and P.D. Pearson. "Some Comments on the Use of Criterion-Referenced Measures in Wide Scale Assessment." *NCME*, (April 1974).

Bloom, B. "Learning for Mastery." *Evaluation Comment*, vol. 1 (1968).

Brown, A. "What Could Be Bad? Some Reflections on the Accountability Movement." *English Journal*, vol. 62, no. 3 (March 1973), pp. 461-63.

Davis, F. Fundamental Factors of Comprehension in Reading. *Psychometrika*, 1944, 9, pp. 185–97.

Guthrie, J. "Models of Reading and Reading Disability." *Journal of Educational Psychology*, vol. 65, no. 1 (August 1973), pp. 9-18.

Karlsen, B., R. Madden, and E. Gardner. Stanford Diagnostic Reading Test (Manual), Level I. New York, New York: Harcourt, Brace and World, 1966.

Klumb, R. "The Effect of Incentives for Teachers on Pupils' Reading Skill Development." IRA Pre-Convention Institute X, 1974.

Minnesota Assessment of Reading, *Reading Performance and Descriptive Results for 11th Graders and Seventeen Year Olds*. November 1973.

NAEP Reading: Summary, May 1972.

Research Triangle Institute. Reading Performance for Seventeen-Year-Olds from the Minnesota Educational Assessment, 1973.

Spearritt, D. "Identification of Subskills of Reading Comprehension by Maximum Likelihood Factor Analysis." *Reading Research Quarterly*, vol. 8 (1972).

Stevenson, J.A. and D.D. Johnson. A Comparison of Children's Decoding and Encoding Responses to Synthetic Words. Unpublished manuscript, 1974.

Terwilliger, J. Some Problems Associated with the Concept of Mastery. Unpublished manuscript, University of Minnesota, 1972.

## Manuscript instructions

RT welcomes manuscripts from readers. Consideration is based largely on originality of ideas, freshness of approach, pertinence of topics and sound scholarship or solid challenge. *Instructions for Authors* tells how to submit a manuscript for consideration by the journal. To obtain a copy, write Journal Editor, International Reading Association, 800 Barksdale Road, Newark, Delaware 19711.