Changing Stage for Students, Teachers and Schools

Michael Lamport Commons

Harvard Medical School

Productivity and therefore income depends on education. But education is suffering world wide. The problems with education are not with the lack of effective pedagogy. It is with the failure of the educational system to use what is known to be extremely effect pedagogy. Here, we present pedagogical ideas that may help educational systems end their poor performance. Two ideas address these pedagogical issues and include two variables that are essential in efforts to improve education including: the actual stages of performance of teaching and learning; and the set of motivational conditions that support increases in the stage of performance to meet task demands. This paper situates the discussion of these variables in the context of a sequence of the minimum behavioral-developmental stages observed in teaching. It supports the benefit of (a) a developmental framework for individualizing instruction and (b) plans for motivating for students, teachers and administrators. Individualized instruction and reinforced successful performances tailored to teachers, students and administrators are two of the general methods introduced. Another is a computerized Reading Teacher based on stacked neural networks, an approach to a behavior-based flexible artificial intelligence. Attending to individual patterns of development on specific tasks along with the use of reinforcement for immediate performance should make it possible for up to 95% of the children to learn to read.

Periodically, agencies of the federal government issue reports on the state of education in the United States (U.S.). For example, in a report issued by the National Center for Education Statistics (NCES), the following quote appears: "While our younger students are making progress on national assessments and are ahead on some international measures the same can not be said at the high school level", Schneider (2006), NCES Commissioner. "U.S. students compare relatively well in reading literacy when compared to their international peers, but they are outperformed in mathematics and science and our 15-year-old students trail behind many of our competitors in math and science literacy." Despite multiple national efforts undertaken from different perspectives over a number of years, education continues to fail many students. As a result, there is a continual need for new ideas to improve teaching and methods to obtain learning and development. The current paper presents a set of new ideas to achieve these aims.

These ideas address issues in pedagogy alone. Nothing else has been shown to be very important. There are two variables that must be examined in improving education. The first is actual stage of teaching performance and of performances of students, administrators and the like. Discussed below are three aspects of stage of performance: minimum stage demands to be successful at teaching material at a given grade; what is normative now; how computer assisted instruction based on stacked neural networks might solve these problems. The second is the set of motivational conditions that support stage change as the hierarchical complexity of problems demand more complex tasks.

The purpose of this paper then is to present a sequence of minimal behavioral developmental stages at which teaching takes place. It also describes the activities engaged in at that specific stage and presents the pedagogical problems with low-stage teaching. Low-stage teaching is supported by two conditions: (a) there are not enough higher stage teachers in the system and (b) teachers that are at a high enough stage are not receiving the support they need for teaching at the appropriate stage. In Table 1, the minimal observed-stage teaching skills of teachers at that stage are listed. Note that these are not the minimal skills needed to teach successfully but a description of what one typically finds.

Problems and Methods to Ameliorate Them

There are serious problems that may occur at the three lowest stages of teaching. For example, at the primary stage, teachers fail to integrate the perspective of the student and their own to allow for negotiated contingency setting and only authority is used. At the concrete stage, people do not benefit from the social norms of the teaching activities. In effect, the teacher just talks "at" the students, rather than "with" them. This describes rigid and unresponsive application of rules given by authorities. At the abstract stage, teachers do not understand what works or not for a particular student or group of students. These are problems with teachers acting at the primary, concrete and abstract stages, where they do not treat students as individuals with individual needs.

Despite the potential for these serious problems to arise at the three lowest stage of teaching, it is entirely possible for teachers to be effective at each stage if the technology they apply is adequate for the teaching task, or if they receive other kinds of support for their teaching activities. Supports may consist of step by step manuals or procedures they are to follow. For example, at the concrete stage, the teacher might only administer computer-aided instruction that automatically adjusts its content for student stage of performance on each task sequence. At the abstract stage, teachers could receive extensive training and support, which may allow them to learn skills to provide empirically-based solutions. Even at some of the higher stages of teaching, some additional support may improve outcomes. For example, at the formal stage, the teacher could be provided with support to generate more effective empirically-based solutions to individual student problems.

Another approach is to increase the stage of teaching of the teachers. There is an overall behavioral acquisition model of how to bring about stage transition (Commons & Richards, 2002). This model proposes a series of transition steps between stages that may occur during the acquisition of the next stage behavior. An understanding of an individual's current transition step influences the choice of intervention to be used.

Table 1. General Stage of Performance Required for Teaching at Different Levels

Stage	Teaching Level	Action	Students' Grade
Primary 7	Teacher's Aides	Follow instructions and imitate modeled behavior	Grade 1-3
Concrete 8	Teachers in early grades of Elementary School K-4	Follow a manual and effectively carry out procedures	Grade 4-6
Abstract 9	Late Elementary School Grades; and Junior High 5-9	Carry out the normative teacher behaviors	Grade 7-10
Formal 10	High School teachers 10-12	Graph student performance and adjust tasks to fit student performance	Grade 11-16
Systematic 11	Four and five-year college professors	See multivariate determinants of student performance	Graduate School
Metasystematic 12	Professors at research universities	Design an entire educational enterprise that works well such as computer-aided instruction	n/a

At step 0, failure of individuals' current strategies of doing things sets the conditions for them to increase their behavioral variability. They try different behaviors. Stage change may take place if new behaviors are modeled, reinforced, or automated as in computer-assisted instruction. At step 1, teachers try a new behavior, often after having seen another teacher do it, or after attending a workshop. At step 2, they alternate between these two behaviors, sometimes using the old one and sometimes using the new one. At step 3, they fit these two behaviors together, much like a 4th grader learning how to fit multiplication and addition together in long multiplication. That new combination of behaviors may be at the next stage if it is defined in terms of two or more lower-stage behaviors, organizes their order of execution, and that order is not arbitrary.

How teachers are selected by schools and states can have the effect of frequently selecting for teachers who perform at the lower stages. Teachers also may not prefer to work with students acting at more than one stage lower than them. As a result, this may limit the stage of individuals choosing to work at particular grade levels as shown in Table 1. For example, by performing at a higher stage, teachers' effectiveness may increase but their level of interest in the teaching activities may decrease because the students' ages and stages do not need higher stage teaching capacities. Stage change, therefore, may lead individuals to leave teaching for better opportunities. Those teachers who have experienced a stage change may not fit in with other teachers, which could also precipitate a change of career. Conversely, performing below the lowest stages described in Table 1 may also lead to teachers leaving the profession. Their performance would be considered a failure by their peers.

Creating an effective educational system

In order to create an effective educational system, there are a few basic requirements. Compared to the system in place, the new system would have to

1. Cost the same or less
2. Utilize current teachers and move teachers or administrators up in stage of performance
3. Require the same or less effort or time by anyone involved.

Such a system would include group and individual contingencies at all levels, as well as reinforcers that matter to the least-motivated performers. This system would also need to include assessment of gains and losses in performance at all levels. Another major contributor in such a system would be automation. This idea of automation would need to include dynamic feedback at all levels, be able to easily accept revisions, and be sensitive to the stage of performance of everyone.

Assessment

The core of assessment is to look at student *changes in performance* on different time scales. It is important to look at how each student performs in each subject. There are two reasonable ways to complete this analysis: (a) chart the behavior as in precision teaching (Graf & Lindsley, 2002; Pennypacker, Gutierrez, & Lindsley, 2003); (b) give both pretests and posttests for every subject every year (this yields difference scores). One might also give stage assessments to each student for every subject, as well as conducting long term follow-ups. One can assess the order of hierarchical complexity for each test item, the level of support to derive the overall task demand (Crone-Todd, 2007). The Task Demand is the order of hierarchical complexity minus the level of support.

When conducting assessments it is important to use an absolute scale. Stage of performance on tasks from a behavioral sequence is an absolute behavioral measurement scale. This is because it is based on the order of hierarchical complexity of the task successfully addressed. The stage achieved in each academic area yields a profile and the average stage divided by the highest stage yields the breath of performance.

Effective Methods

There are some things that do not seem to affect assessments much, if at all. One is school structure, which accounts for a minuscule amount of change (Richards, 2002, personal communication). The curriculum is a second variable that also produces very small amounts of change. Two things (Walberg, 1984) that do seem to contribute are individualized curriculum and the motivational contingencies (Flora, 2004; Flora & Flora, 1999). Herbert Walberg (1984), in his review of the research literature, concluded unambiguously that Skinnerian reinforcement or reward for correct performance has the largest overall average effect: 1.17 standard deviations (p. 23). A review of Walberg's results by Bloom (1984)

found that only one-to-one tutoring had a greater effect on student achievement than reinforcement. Note that one-to-one tutoring probably involved reinforcement as well. The individualization of what a student works on should be based on stage and step-in-stage-transition sequence where the student is performing. The motivational contingencies that are effective for individual teachers themselves are part of having contingencies at all levels. In order to be maximally effective, contingencies should apply at all levels and be based on behavior. The levels would be the students, teachers, departments, schools, districts, states, and countries.

The motivation and individualization of material while paying attention to stage are matters of concern to the teaching process. This motivational analysis grows out of our experiences at Dubnoff School for Educational Therapy in North Hollywood, CA, Project Giant Step in the community school Ocean Hill Brownsville in Brooklyn, NY (Littky, Commons, Goodman & Shulman, 1959), and Morse School in Cambridge, MA (Commons, Grotzer, & Davidson, in preparation). Those contingencies were adapted from recess period games, which the children play by themselves without supervision. Recess is the most preferred part of the school day. Therefore it was assumed that the activities and games within it were highly reinforcing. Note that they play the same games all the time (Goodman, 2006, personal communication. Competition is another variable that can be used to produce large amounts of learning. This occurs by using group contingencies based on everyone's individual performance. At the same time, each individual's performance adds to the group total, creating both individual and additive effects to learning. This contingency provides an equal opportunity and equal duty to contribute, which closely mimics the real world and includes a natural weighting system.

In our own research, we have used such group-individual contingencies to produce marked improvement in teaching and learning. At the impoverished Ocean Hill Brownsville School, Project Giant Step produced an improvement in reading performance from an average of 0.4 years per year to 1.8 years per year in the student participants. At the Morse School in Cambridge, we hoped to facilitate a progression from the concrete stage to the formal stage in the 5th and 6th grade participants. After 16 sets of interventions, 75% of the students performed at the formal stage, which represented an increase of 55% from their pre-intervention ability. There are some contingencies that do not work as well as the aforementioned ones and the examples given previously. One is the traditional use of individual competition, which does not develop cooperation and fails to get children or adults to teach each other. A second contingency that often does not work well is using groups without attending to individual performance (Cooperative Learning). This contingency is common in schools, but typically results in one or two people doing all of the work. Thus, there becomes no reason for the students engaging in social loafing to perform; their learning of the material is severely compromised. Another example is using groups, but weighting the contributions of the top five to ten performers. This is often done in sport settings. Using this contingency often provides a disincentive to the weak performers, while at the same time labeling them negatively and making them feel as if they do not contribute at all.

Another important aspect of facilitating a positive change in performance for students, teachers, departments, schools, etc., is simply observing when a negative performance is occurring. When this is recognized, someone can be assigned to deal with the problems that are occurring by providing modeling and training in an appropriate fashion, as well as initiating supervision to monitor progress. Once a positive change is initiated, it is also helpful to introduce positive payoffs, which produces acceleration in the positive change.

One type of instruction that has been shown by behavior analysis to produce marked improvement is individualized instruction. However, the problem is that using individual teachers and aids to make decisions as to what should be presented often fails. This is because such decisions require formal stage performance, which is typically not reached by lower-stage performing teachers. Besides the lack of formal stage performance in the teachers, there is not often the time to individually teach all students. Yet, it is critical to not lose sight of individualized instruction, as it does occur and is useful for changing stage. When analyzing individual instruction, two variables typically make most of the difference. The first are the stage demands of the task on which the person is working. For example, in learning to read, stage demand is crucial because there are not only stage problems, but also sequencing within stage problems. The second important variable is the current level success on a given task. For short term assessment, one can use the celeration chart (Graf & Lindsley, 2002). For long term assessment, it is more effective to use standardized tests as a way to gauge overall success of the program and of the student.

Hierarchical Analysis

Automatic ways of detecting change in performance exist, such as using computer assisted instruction (CAI) and computer management. By using CAI, it is possible for the computer to continually assess performance. With that assessment, the stage requirements and other sequencing issues can be automatically adjusted. This allows for individualized adjustable, and differential, reinforcement contingencies to be put into place. Another strength of using CAI is that its programs and procedures can be precisely described and constantly improved on the basis of student performance.

In order to account for the impact of various levels of input, it would be important to use a Hierarchical Linear Model. This is because we have hierarchical or nested data structures, which consist of the following pattern:

A country consists of many states
A state consists of many school districts
A district consists of many schools
A school consists of many departments
A department consists of many teachers
A teacher applies many contingencies to the students
A teacher applies a varying degree of mismatch between task order and stage of the student
A group of students in a class have variously performing peers and parents

When an understanding of impact at each level is achieved, one can begin to find out where to intervene and what interventions work. Over time, with the use of scores that represent changes in stage, one can discover what role each level plays in producing the change; and whether it is helping or hindering the change needed to progress within and between stages.

Another benefit of using hierarchical analysis is that it can be used to see what to check when interventions fail. When the failure occurs, hierarchical analysis can determine if it is due to a mismatch in the stage of the student performance and the task at hand, if there is a lack of reinforcement contingencies to promote positive change, or the existence of reinforcement contingencies to maintain the status quo. Unfortunately, a hierarchical analysis will always have to depend on incomplete data, but even this incomplete data ends up being quite robust. During the analysis, one can start with any part of the data and then build out to other levels or analyze other players at each level. This allows for increasingly helpful comparisons and greater accountability.

By finding the hierarchical complexity of the items, it might be possible to predict the Rasch (1980) scores of both the items and of the participants. This would allow for pretest-posttest difference

scores to be analyzed for test items, tests, or groups of tests. Difference scores are important for making assessments.

Last, very little research has gone into applying stacked neural networksh [1] to teaching problems. An example follows as to how stacked neural networks could be applied to the most difficult teaching task, the teaching of reading. The rest of this paper will address the usefulness of such networks in terms of teaching reading.

Using Stacked Neural Networks in Teaching Reading

Because tutoring is the single most effective form of teaching (Walberg, 1984), especially for reading, it is important to conceive a means for providing every child with his or her own tutor without requiring a roomful of tutors or a tutoring service. In the future, reading may be extremely efficiently taught using stacked neural networksh without the need for teachers or tutors. At present, the reason teachers are necessary is that they can tell if the student is reading correctly. But it is possible to have this function done by trained stacked neural networks. To teach reading, the following stacked neural networks are presented on a computer. Stacked neural networks simulate some of the functions of the nervous system. Each level corresponds to a given layer in the organism and at the same time a given stage of development. The outputs from one layer are fed into the next layer above even though all the layers are interconnected to some extent. Each stacked neural network level consists of multiple neural networks. Each layer is trained by having it interact with the real world rather than with a programmer. It learns from consequences produced by its choices. For the example presented here, we restrict this to learning to read letters and words out loud. This might be called a behavior-based flexible artificial intelligence.

We begin at the level of presenting stimulus material. First, a letter is displayed on the screen. The stacked neural network says the name of the letter. Out loud, it asks the participant to say it. The stacked neural network is set to discriminate varying degrees of proximity to what the participant utters. The stacked neural network is able to shape what the participant says when presented. It starts with rather large deviations. If they are in the bounds, they are scored as correct, and the participant receives a point. If not, they receive nothing and go on to the next item.

An artificial intelligence (AI) program that teaches reading may evolve because of programmer interventions. It may even learn in an interactive way in real time from what the participants do. But such programs are relatively inflexible. What they accept is usually predefined. AI programs are relatively insensitive to variations that are not in their database. The users are the ones who have to adjust their search words to get a program that teaches reading to find what the users want. Often there is failure. In contrast, each of the stacked neural networks incorporate the work describing the previous networks which are stacked neural network in and of themselves. In each of the levels, stacked neural networks have different goals. Many work together because they work on the output of the previous layer and tasks partially accomplished at that layer. The stacked neural network of interest here is the Teaching Reading Programh.[2]

To teach reading, there are at least two broad, compatible approaches. One is to teach writing at the same time so that the computer receives letters keyed in by the participant. The other way is for the participant to speak and the computer to recognize what has been said. Neural networks have been applied to both character and voice recognition. It is not difficult to train neural networks to recognize a few words. The problem is trying to train them to learn hundreds of thousands of words and ensuring that the strings of words make sense. Stacked neural networks approach the problem of editing at different levels, making a solution tractable. In all cases, a large number of paragraphs written by a large number of writers would be fed to the stacked neural networks. This is standard procedure in character recognition and voice recognition. What will be different is that meaning will be developed in the higher levels of the stacked neural networks, making the choice of words more accurate and the spelling more correct. In general, because each layer will provide output to the user, the user can also provide input to the next layer. This concept is introduced in the following examples, which present only the first three layers of such a network.

Stage and Layer 1: Character and Phoneme Identification

The first layer of neural networks recognizes phonemes or characters. One would create 256 neural networks, each tuned to recognize just one character or phoneme. The phonemes and characters would produce letters and series of letters. If there are identification conflicts, such as between “g” and “q” or “t” and “I,” the next higher level network would process both possibilities. There is also misreading by the student reader. The neural networks would provide feedback internally to the program regarding whether these vocalizations of letters was correct or not, and if not they would scale how far off the response is from the correct one. This information would be used to decide whether to give a point for the vocalization. The goal would be to keep the percent of reinforcement at about 85 to 90%. Therefore, there would be a slight tightening of the criteria over trials for a given set of characters. These character identification networks would also detect spaces and punctuation that mark end of words, phrases and sentences as well as pauses in speech. The vision is that the output would then be fed to a spell Teacher network layer. One form of training for each neural network at this stage would consist of providing a large number of representations of the same character in different fonts and in different handwritten form or the same vocalized phoneme from different people and presenting cases that might be confused, such as similar characters or sounds. Another would be to provide a large number of written documents with words containing the letter or representation of the phonemes for each neural network and presenting distracters as well. Order is preserved, and characters and symbols are passed forward to the next layer.

Stage and Layer 2: The Reading Teacher Word Identifier

The Reading Teacher will recognize single spoken or written words built out of the character strings passed from the character identification network. It will also recognize the new reader’s utterances. Ambiguity in character identification and phoneme recognition in Stage 1 creates multiple possible character strings, now considered as words. At this point there is no meaning to a word. The neural networks begin with lists of characters and lists of possible words. They learn to identify the unique word sought, given the context of the word’s usage. Note that identifying the correct word (and spelling) is not error-free at this point because there are not enough constraints, such as context. If the neural network finds a match between what was presented visually with what was said, it would display that the person earned a point. The neural network can ask the reader if this is what they are saying in response to what the reader said. Training consists of presenting many documents or speech samples composed of the same words, distracter words, and slightly different words. It could pass on good word possibilities both to the user and the next layer.

Stage and Layer 3: A Reading Teacher

Common errors are made in English usage and detection of what has been said out loud. Examples include misuse of articles “a”

[1] Michael Lamport Commons and Mitzi Sturgeon White

[2] Michael Lamport Commons with Leonard Sidney Miller

and "the" and words, such as "real good" versus "really good." Other mistakes arise from homonyms: "reign" versus "rain," or "to" versus "two" or "too," "principle" versus "principal," or "who" versus "whom." etc. A word usage Teacher would teach correct word pronunciation. The neural network would be trained by presenting words with the same or similar sounds from the written letters, and output could be fed to the next layer.

Conclusion

Attending to individual patterns of development on specific tasks along with the use of reinforcement for immediate performance should make it possible for up to 95% of the children to learn to read. Although stage of development on individual tasks is assessable, it is almost never done. Placing children on teams and reinforcing individual performance immediately is also almost never done. The problem is not how to teach reading as much as how to get the educational system to use what is effective in teaching reading. With the invention of stacked neural networks, there will be a way to make the process of assessment, individualization and motivation available to all students independent of the quality and training of teachers.

References

Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher, 13*(6), 4-16.

Commons, M. L., Grotzer, T. A., & Davidson, M. N. (In preparation). *The sufficiency of reinforcing problem solutions for transition to formal operations: An examination of Piaget's equilibration theory of stage change.*

Commons, M. L., & Richards, F. A. (2002). Organizing components into combinations: How stage transition works. *Journal of Adult Development, 9*(3), 159-177.

Crone-Todd, D. E. (2007). Assessment of thinking in adult learners. *Behavioral Development Bulletin, 13,* Spring, 41-44.

Flora, S. R. (2004). *The power of reinforcement*. Albany, NY: State University of New York Press.

Flora, S. R., & Flora, D. B. (1999). Effects of extrinsic reinforcement for reading during childhood on reported reading habits of college of students. *The Psychological Report,* 49: 3B14.

Graf, S, & Lindsley, O. (2002). *Standard Celeration Charting*. Youngstown, OH: Graf Implements. Available from Graf 1. Implements; e-mail: Zerobros@aol.com

Littky, D., Commons, M. L., Goodman, S. A., & Shulman, F. (1959). *Project Giant Step*. Ford Foundation Report.

Pennypacker, H. S., Gutierrez, A., & Lindsley, O. R. (2003). *Handbook of the Standard Celeration Chart*. Cambridge, MA: Cambridge Center for Behavioral Studies.

Schneider, M. (2006). *The Condition of Education.* Retrieved March 5, 2007 from http://nces.ed.gov/programs/coe/press/highlights.asp

Walberg, H. J. (1984). Improving the productivity of America=s schools. *Educational Leadership*, 41, May, 19B27.