Prof.dr. Christine Espin

# Realistic and Relentless

Using data-based decision making to build effective programs for secondary-school students with learning and behavior difficulties.


Universiteit Leiden

# Realistic and Relentless.

Using data-based decision making to build effective programs for secondary-school students with learning and behavior difficulties.

Oratie uitgesproken door


## prof.dr. Christine Espin


bij de aanvaarding van het ambt van bijzonder hoogleraar op het gebied van Diagnostiek door Leerkrachten van Leer- en Gedragsproblemen bij Adolescenten aan de Universiteit Leiden
vanwege het Leids Universiteits Fonds
op maandag 10 mei 2010.


**Universiteit Leiden**

*Mijnheer de Rector Magnificus, Mijnheer de decaan, leden van het Curatorium van deze leerstoel, zeer gewaardeerde toehoorders,*

One of the advantages of moving to a different country is that it allows you to view things as an "outsider" - for both your new country and your old country. Although making gross comparisons is always a bit risky, one does notice differences that seem to be country specific. For example, I notice that Dutch cheese is somewhat better than U.S. American cheese - or as my Dutch husband says, "Nederlanders hebben kaas - Amerikanen hebben plastic". And Americans in the U.S. tend to be more open and outgoing with strangers than are the Dutch. For example, you are probably familiar with the well-known American greetings, "*Hi! How are you?!*" or "*Hi! Where are you from?*" - greetings which Americans find perfectly appropriate for people they have never seen before.

Some between-country differences are of a more serious nature, of course, and can have greater consequences for the citizens of those countries. It is one of these differences that provides the catalyst for my talk today. I refer to this difference as the "*Realistic vs. Relentless*" difference, or in Dutch, "*Realistisch vs. Volhardend*" *verschil*. To illustrate this difference, I ask you to consider the concept of the *normal curve.*

Many of my academic colleagues are quite familiar with this concept, but for others an explanation may be helpful. The normal curve, or in Dutch, *de normaalverdeling,* is important to research in education and psychology. Much of our research is based on the assumption that human characteristics fall along a *normal curve* much like that depicted in Figure 1. To illustrate, consider the human characteristic of height. If we were to know the height of every person in the world, we might assume that the values would distribute themselves in the shape seen in the figure, where the mean height is represented by the line in the middle of the curve, and the majority of the population falls within one standard deviation - that is one block - above or below the mean. The further one moves away from the mean, the smaller the percentage of the population in those blocks; thus, only a small percentage of people would be assumed to be much, much taller or much, much smaller than average. In this room filled with mostly Dutch people - who as a people tend to be tall - many people would fall above the mean in the distribution. (An exception would be my mother and sisters who are of Italian descent and would probably fall 1 1/2 to 2 standard deviations below the mean!)

So how does the normal curve relate to the topic of *realistic vs. relentless*? It seems to me that one fundamental difference between the U.S. and Dutch educational systems is the *acceptance of the normal curve*. From my observations, the Dutch seem to be generally accepting of, or *realistic* about, the existence of the normal curve. Applied to education, there seems to be an acceptance of, or at least an implicit belief that, every child comes to school with a different inherent learning capacity. The goal of the educational system is to identify the child's learning capacity, and then match the child to an appropriate educational program. As a result, there are a multitude of different types and levels of education, and much time, energy, and attention is devoted to matching learner to program.

What about the U.S. American view of the normal curve and the resulting approach to education? If you examine Figure 2, you see my depiction of the American view of the "normal curve." What is immediately obvious is that this is not a normal curve at all. In this distributional curve, everyone is average or above (a statistical impossibility).

From my observations, in the U.S., the Americans seem to be generally *not* accepting of, *not* realistic about, the existence of the normal curve. Applied to education, there seems to be a belief that, although every child may come to school with a different inherent learning capacity, this capacity should not be seen as an impediment - or in Dutch, a *belemmering*

- to having the child achieve average or above. The goal of the educational system is to be *relentless* - that is, to ignore the child's inherent capacity and move the child up the achievement continuum as far as possible. This view is reflected in the educational rhetoric of America, where one often finds phrases such as "Every child can learn," and "No Child Left Behind". As a result of the relentless view, types and levels of education are as similar as possible for all children, and much time, energy, and attention is devoted to leaving all choices open to all children for as long as possible.

**Is one approach better than another?**
Such depictions of country-based educational differences are, of course, over-simplifications, but they do serve to illustrate fundamental differences in the general approaches to the education of children. The logical question that arises is, "Which approach is better?" Is it better to be "realistic" - *realistisch* - or to be "relentless" - *volhardend* - in the education of our children? I believe that the answer to that question can be found in a quote from the famous Dutch "philosopher", Johann Cruijff: "Ieder nadeel heb zijn voordeel", which loosely translated means, "Every disadvantage has their advantage".



My argument would be that each approach has advantages and disadvantages. The advantage of a *realistic* approach to education is that the educational program is matched to the capabilities of the students, and therefore should be of high quality for those students. The disadvantage is that there may be little room for flexibility in the system, should the original matching be incorrect, or should students want to try a more difficult program of study. The lack of flexibility arises from the assumption that students are unlikely to do better than the level that has been selected for them. That is to say, there may be little expectation that students will "rise above" their inherent capabilities.

The advantage of the *relentless approach* is that a multitude of opportunities remain open to students for a long period of time. There is an expectation that, at any time, with hard work and effort on the part of educators, parents, and students, the students might "rise above" their inherent capabilities to succeed beyond expectations. The disadvantage of the relentless approach is that expectations for students and schools may be unrealistically high, and high quality programming may be sacrificed to achieve flexibility and choice. Unrealistic expectations may set students up for failure, or lead to programs where students "succeed" because they graduate, but where they learn little or do not attain the skills they need to be successful after graduation.

Is the approach taken to education really that important? I would guess that for many students - maybe for most - the general approach taken to education has limited impact. That is to say, many students eventually find their way into programs that suit their interests and capabilities, and they learn enough to achieve success, regardless of the educational approach. However, for students with mild learning and behavioral difficulties - that is students with the labels dyslexia or dyscalculia or Attention Deficit Hyperactive Disorder (ADHD) - for these students, the selected approach may have a major impact. In the Netherlands, such students may be denied access to a level of education where they *could* succeed if given appropriate supports and interventions. In the United States, such students may be denied a high quality educational program because they *must* be placed in a program and curriculum designed for students without disabilities.

Is there a way out of the *realistic vs. relentless* dilemma? Can we build educational programs that are both *realistic and relentless* for students with special needs, or are the two approaches orthogonal to each other; that is, must we have one or the other?

I believe that it is possible to be *both* realistic and relentless in our educational programming for students with special needs - and that one important step to achieving a realistic and relentless program is for educators to become effective data-based decision makers.

**Data-based Decision Making**

What is *data-based decision making?* In a data-based decision making approach, educators use student performance and progress data to make informed educational decisions for students at risk. Figure 3 illustrates one particular data-based decision-making approach, referred to as Curriculum-based Measurement or CBM. This data-based decision-making approach is the focus of my research.

The graph in the handout represents the performance level and rate of growth across a school year in reading for a student named Tom. At the beginning of the school year, it is obvious that Tom is performing far below his peers. Whereas his peers have an average score of 140, Tom has a score of 41. However, the fact that Tom performs at a level below his peers does not necessarily mean he will not profit from the same instruction as his peers; thus, Tom's performance is sampled weekly and the scores are graphed. After 6 weeks, a line of best fit is drawn through the data to represent Tom's growth. It is easy to see that Tom is not profiting from typical instruction. He is not improving in reading.

In response to the data, a change is made in Tom's instructional program, and data continue to be collected weekly and graphed. After 6 weeks, a line of best fit is again drawn through the data. As you can see, Tom continues to struggle. Something more intensive is needed. At this point, additional testing may be done to further understand the nature of Tom's reading difficulties, and a specialized program may be designed involving the use of different curricula, materials, or instructional approaches.

A specific goal is set for Tom (represented by the solid diagonal line), and data are collected weekly. As with the previous phases, after 6 to 7 weeks, the data are evaluated to determine the effectiveness of the program, and changes or modifications are made when Tom's progress is less than expected. In this example, Tom profits from the more intensive, specialized instructional program; that is, his rate of growth is steeper than the expected rate of growth represented by the goal line.

As illustrated in the Tom example, the goal of data-based problem solving is to be relentless - yet realistic - in educational programming for students. The fact that Tom begins far behind his peers does not mean he will not learn under typical instructional conditions. Only when both his *performance and progress* data reveal that he is not succeeding, is consideration given to more specialized and intensive programming. Within the specialized program, data continue to be collected and evaluated to determine whether the program is successful, or whether modifications in the program are needed. You will notice that in a data-based problem-solving approach, diagnosis of the disability does not drive educational decision-making - student performance and progress drives educational decision-making.



**Research on the development of data-based decision making**

As you examine the data-based decision making graph, several questions may arise, such as (1) What are the data represented on the graph? (2) How often must the data be collected? (3) How trustworthy are these data? (4) How practical is it to collect such data weekly? (5) How are the expected rates of growth determined? (6) Do practitioners actually use the data to make educational decisions? (7) From what instructional alternatives do practitioners make choices and how do they make these choices?

In 1977, a research program under the direction of Stanley Deno was launched to address some of these questions (see Deno, 1985). This research focused on the development of measures for elementary-school children, or in Dutch, *basisschoolkinderen,* in reading, writing, spelling, and later, math (see Marston, 1989 for a review).

Among the many important contributions of that initial research program was the conceptual approach used to select

or create progress measures, an approach we use in our research today. Specifically, the measures used as a part of CBM must meet both technical and practical requirements (Deno, 1985). Technically speaking, the measures must be valid and reliable if they are to be used to represent student performance and progress, and to guide instructional decision-making. Practically speaking, the measures must be repeatable, simple, efficient, and inexpensive if they are to be administered by educators on a frequent basis, for example once a week. They must also be easy to understand and implement, and must result in practically important outcomes if they are to be useful in educational settings.

Combining the technical and practical considerations leads to the concept of the development of a performance and progress *indicator*. Similar to a thermometer, CBM measures are designed to be *indicators* of *students' academic health* (Deno, 1985). That is, they are not designed to measure specific aspects of learning, or to provide feedback about what to teach. Instead the measures are designed reflect whether what is being taught is leading to improvements in the skill area. For example, in reading, the desire is NOT to have separate measures for word decoding, fluency, vocabulary, or comprehension, but to have one, brief *indicator* that reflects performance and progress in reading in all of these areas; that is to have a global indicator of reading performance and progress.



To examine the validity of potential CBM indicators, the concept of nomological net, as described by Cronbach and Meehl (1955) in their classic paper on validity, is employed. In this conceptualization, the validity of the indicator is determined by examining the pattern of the relations between the indicator and other measures of performance in that area. For example, the validity of a CBM measure in reading (e.g., the number of words read aloud correctly in 1 minute) is determined by examining the relations between that indicator and other measures of reading performance, including performance on standardized achievement tests, performance on reading comprehension measures, the age of the student, teacher judgment, the students' special education status, and so on. Research conducted at the elementary-school level supports the hypothesis that a 1-minute reading aloud measure is a valid indicator of general reading performance. For example, correlations between reading aloud and performance on standardized achievement tests in reading typically range from .60 to .80, and reliability coefficients typically are above .80 (Marston, 1989; Wayman, Wallace, Wiley, Tichá & Espin 2007).

The pattern of relations between an indicator and other measures is an important step to establishing the validity of a measure, but it is only the first step. It is also important to examine the outcomes associated with implementation of the measure; that is, to ask questions such as "What are the effects of CBM progress monitoring on teacher instruction and student achievement?". Such an approach reflects the unified concept of validity as described by Messick (1989a, b). In this conceptualization, validity is not a property of the measure itself, but a reflection of what occurs when the measure is implemented, Specific to CBM, a part of the validity question is what are the consequences of progress monitoring on student achievement, teacher instruction, parental perceptions, etc.

**Extension of Data-based Problem Solving to the Secondary-school Level**

When I began my research career in 1990, there was very little research on the development of CBM measures or on data-based decision making at the secondary-school level. I wanted to explore the development of CBM progress measures for secondary-school students; however, I was immediately confronted with two challenges. The first was *what* should be monitored. The second was *how much improvement* to expect.

With regard to *what should be monitored*, the specific question was "*What is or what should be the curriculum for secondary-school students with learning difficulties?*". Defining the

curriculum seemed, on the surface, to be straight-forward. There was more or less a set curriculum in the United States at that time. For example, students in 8th grade - the second year of secondary school - usually studied subjects such as algebra, life sciences, English, American history, world geography, and sometimes a foreign language such as Spanish. However, the question of curriculum centered not on the content of the established curriculum, but on the extent to which that established curriculum met the needs of students with learning difficulties. For example, should students with severe reading difficulties receive something in addition to, or instead of, the established curriculum? Specifically, should they continue to receive specialized, intensive reading instruction?

The second challenge was related to the first, and addressed the question of *how much improvement to* expect? If students continued to receive reading instruction, how much improvement in reading should be expected? How much did students *need* to improve to be successful following completion of secondary school? And would the amount of improvement justify the time, effort, and resources needed to effect such improvements?

In the 1990s, there was a fairly clear approach to the education of secondary-school students with learning difficulties in the United States. Generally speaking, students with learning difficulties received the same curriculum content as other students, and special instruction was geared toward helping the students earn passing grades – *voldoendes* - so they could graduate from high school. There was little or no attention devoted to basic reading and writing instruction. The general view was, "If they haven't learned it by now, they never will". In this atmosphere, there was no need for a system of progress monitoring in reading or writing for secondary-school students.

This state of affairs began to change in the mid-1990s, when many U.S. states began to enact state standards tests in skill areas such as reading and writing. These changes were related to a standards-based reform movement that eventually culminated in the passage of the *No Child Left Behind Act* (U.S. Department of Education, 2002). In many states, students were required to pass these tests in order to graduate from high school and in many states, a substantial number of students were failing the tests. For example, in the state of Minnesota where I was living, in the first two years that the state reading test was given, approximately 40% of 8th graders failed the test. Suddenly, schools, teachers, and parents were very interested in basic reading and writing instruction for secondary-school students - and, consequently, in progress monitoring in reading and writing. It was probably not coincidence that at about that time we were able to secure federal funding for the development of progress monitoring measures in reading and writing for secondary-school students, and were able to find resources to conduct similar research in content-area learning.



There were two phases to our initial research program. In the first phase, we focused on the development of measures for progress monitoring, in the second phase, on the effects of implementation. I would like to illustrate the line and logic of research by presenting some of our findings in the area of reading. We conducted parallel lines of research in writing (Espin, Scierka, Skare & Halverson, 1999; Espin, Skare, Shin, Deno, Robinson & Brenner, 2000; Espin, De La Paz, Scierka & Roelofs, 2005; Weissenburger & Espin, 2005; Espin, Wallace, Campbell, Lembke, Long & Tichá, 2008), and content-area learning (Espin & Deno, 1993; Espin & Foegen, 1996; Espin, Busch, Shin & Kruschwitz, 2001; Espin, Shin & Busch, 2005).

**First Phase of Research: Development of progress measures**

In the first phase of research, we examined the validity and reliability of two potential indicators of performance and progress in reading (Tichá, Espin & Wayman, 2009; Espin, Wallace, Lembke, Campbell & Long, 2010): reading aloud and maze selection. We created reading aloud and maze selection measures from newspaper articles. For the reading aloud

measures, students read aloud for 1, 2, or 3 minutes, and the number of words read correctly was counted. For the maze selection measures, every 7th word was deleted and replaced with a 3-option multiple-choice item. Students read through the text and selected answers as they read. For each type of measure, we examined different time frames - for example, 1, 2 and 3 minutes of reading - and different scoring approaches - for example, scoring correct only vs. scoring correct minus incorrect. We also looked at the characteristics of the measures as both performance and progress measures; that is, we examined whether the measures reflected students' level performance compared to peers, and whether the measures reflected progress or growth over time.

With regard to ***performance,*** our results supported the hypothesis that both reading aloud and maze selection were good indicators of a student's level of reading compared to his or her peers (Tichá et al., 2009; Espin et al., 2010). Correlations between the CBM measures and performance on a state reading test and a standardized achievement test in reading ranged from .75 to .89. We found few differences related to time frame or scoring procedures.

With regard to ***progress***, we found differences in the characteristics of the measures for reflecting growth. Across two different studies (Tichá et al., 2009; Espin et al., 2010), we found that reading aloud reflected little to no growth over time. This pattern of results was true regardless of time frame or scoring procedures. In contrast, maze selection produced relatively stable, linear growth rates, and these growth rates were related to performance on the state reading test, and change in performance on a standardized achievement test. We speculated that on the reading aloud measure, student reached a natural level of fluency that served to differentiate them in terms of reading skills but did not reflect change over time. We further speculated that maze selection was sensitive to growth because it reflected a broader range of reading skills, including fluency, word recognition, vocabulary, and comprehension.

In 1990, we received federal funding for the Research Institute on Progress Monitoring. The Institute was co-directed by Dr. Teri Wallace and myself, and involved 7 colleagues from the Universities of Minnesota, Iowa, and Missouri, working collaboratively on the development of progress measures in reading, writing, and mathematics for children ages 4 to 18, both with and without disabilities. In our work at the Institute, we examined the technical adequacy of CBM measures in reading both for following growth *across* as well as *within* academic years. In reading, our results revealed that a 3-minute maze task created from a 4th-grade reading passage (that is an AVI level 7 passage), could be used to follow growth across school years, specifically, from grades 3 to 10 --- that is, Groep 5 to the second year of middelbare school (Espin, Wallace, Tichá, Wayman, Wiley & Long, 2006).

**Second Phase of Research: Implementation of progress measurement**

In the first phase of our research program, results had provided support for the maze-selection measure as a valid and reliable indicator of performance and progress for secondary-school students. In the second phase of our research program we turned our attention to the effects of progress monitoring implementation. We examined whether there were educationally significant outcomes associated with implementation of the maze selection for monitoring progress.

In our first study, we examined the effects of implementation on teacher instruction and student performance (Espin, Wallace, Long, Lembke, Campbell & Tichá, 2003). We hypothesized that if teachers collected progress data to evaluate the effects of their instructional programs on student performance, they would build more effective instructional plans in response to the data, and, in turn, students would achieve more. In a within-teacher design, we randomly assigned students to a teacher-monitoring vs. researcher-monitoring condition. In the teacher-monitoring condition, teachers monitored and graphed student performance weekly

for the entire school year in reading, and used the graphs to evaluate the effects of their instruction on student growth. In the researcher-monitoring condition, researchers monitored the students and did not show the graphs to the teachers. Results revealed that students in the teacher-monitoring condition grew significantly more on the progress measures over the course of the study than students in the researcher-monitoring condition, but these differences did not translate into differences in scores on the state reading test. In addition, we found no observable differences in teachers' instruction for the students in the two groups.

In a subsequent study, we examined the effects of student participation in progress monitoring, randomly assigning students within teacher to a graph sharing vs. no-graph sharing condition (Wallace, Espin, Tichá, Wayman, Wiley & Long, 2005). Teachers monitored progress weekly for all students, but shared the graphs with only half of the students. We hypothesized that showing students their progress graphs would motivate them to do better. Despite our high hopes for this study, we found virtually no effects associated with student participation in monitoring. Following the study we conducted focus groups, and asked students whether they liked seeing their progress graphs. Their answer was, (shoulder shrug), "Yeah, it was OK". We asked if they would like to continue to see their progress graphs, and they answered, (shoulder shrug), "Yeah, that would be OK". We asked if they thought it was a good idea to share their progress graphs with their parents. They sat up straight, and answered, "No! Absolutely not! Our parents would just tell us to work harder!".

Despite this reaction, in the following year (with a different group of students), we examined the effects of sharing data with parents (Campbell, Wallace, Lembke & Espin, 2005). We hypothesized that graphs would be a simple and efficient way to communicate progress information to parents. Participants were parents of at-risk high school students. The students were enrolled in a 6-week summer school program. We monitored students' performance in reading and created a progress graph for each student. At the end of the study, all parents received detailed narrative reports describing the student's progress during the 6 weeks. We randomly assigned parents to a graph vs. no-graph condition. Half of the parents received the summer-school reports with a progress graph, the other half without. We then asked parents to complete a questionnaire about their child's performance during the summer school program. Results revealed that parents who received progress graphs were more positive and more accurate in their judgments about their child's progress than parents who did not see the graph.

To recap, in the initial two phases of our research program, our research revealed that the maze-selection measure had reasonably good reliability and validity both as an indicator of performance and progress in reading, and that there were *some* positive effects associated with implementation of progress measures for teachers and parents.



One of the puzzling things about this early research was the modest effect associated with teacher implementation of progress monitoring. That is, although we found significant effects associated with progress monitoring on growth, the effects were small in magnitude and did not transfer to the state reading test. To explore this point further, we examined the teachers' instructional plans for the students who had been monitored, and found that teachers seemed to not *use* the data to make instructional decisions. That is, although the teachers reliably collected, scored, and graphed the data, they did not change instruction in response to the data - this despite fact that the research team reviewed graphs on a regular basis and prompted the teachers to make changes when students were not progressing.

This problem of teacher data use was not new. It had been observed in previous CBM research (see Stecker, Fuchs & Fuchs, 2005 for a review), and in research with other formative

assessment systems (e.g., Black & Wiliam, 1988; Black & Wiliam, 2005; Tillema, 2009; Tillema & Smith, 2009).

I found the problem of teacher data use both discouraging and fascinating. It seemed as though CBM progress monitoring might prove to be a reasonably good tool for depicting student progress, but if teachers did not - or could not - use the data, the tool would be useless. I could not help but wonder *why* teachers did not use the data. Was it that teachers were limited in their ability to interpret and use data, or was there something more fundamental at work - something related to human's general ability to use data to make decisions?

Thus, in the third, and most recent phase of my research, I have turned my attention to teachers' use of data for decision-making. I began this phase of my research by reading the literature on decision-making in general, and teachers' decision-making in particular. Before going on, I would like to mention that others here in Leiden, such as Jan van Driel and Nico Verloop from ICLON, and Harm Tillema and former Leiden colleague, Mien Segers, from Onderwijsstudies, do related work in areas of teacher thinking and the use of assessment data to inform instruction.



**Third phase of research: Teachers' use of data for decision-making**

As I read the decision-making literature, it became immediately clear that, in general, human beings are not very good at using data to make decisions. More specifically, we humans use data in only a limited fashion for decision-making. The Nobel Prize Winner, Herbert Simon, whose work was informed by the eminent Dutch psychologist, Adriaan de Groot (De Groot 1946; 1965; Vicente & De Groot, 1990), referred to this phenomenon as "bounded rationality" (Simon, 1990). Simon argued that humans are constrained in their decision-making capabilities by invariants in human behavior such as limited short-term memory, recognition time, and reaction time. At the same time, humans are faced with a highly complex world in which decisions must continuously be made. To deal with the problem of limited capabilities in a complex world, humans adopt strategies to simplify their decision-making, such as using recognition, selective or heuristic searching for solutions, and serial pattern recognition. In more recent years, human decision-making has been characterized as "fast and frugal", (Todd, 2007; Perkins, 2009), implying that we humans use just enough data to come to a decision that is good enough, and then move on.

Time does not permit me to explore all of ramifications of the decision-making literature on the study of teachers' use of data, however, one point has become quite clear. The step from data collection and graphing to data use and decision-making is not a small, inconsequential step for teachers. It is a step to be studied and understood. That is to say, it is not enough to develop a reliable and valid progress-monitoring system - it also important to examine the processes involved in teachers' use of that system, specifically their use and understanding of progress data, and their selection of when and how to change instruction in response to such data.

It is just such a study that I recently conducted in collaboration with my former University of Minnesota colleagues, Stanley Deno, Kristen McMaster, and Miya Wayman, and my Leiden colleague, Mark de Rooij (Espin, Wayman, McMaster, Deno & De Rooij, 2010). We asked teachers to examine CBM progress monitoring graphs and complete think-alouds to describe what they saw. After the data were collected, we asked experts to rate the knowledge level of the teachers with regards to CBM. We then examined differences in the think-alouds of teachers' rated more and less knowledgeable. Results revealed that more knowledgeable teachers described the graphs in a more accurate, systematic, and cohesive manner than less knowledgeable teachers. Further, the more knowledgeable teachers described the graphs in a sequence similar to the sequence in which the data would be used to inform instruction - from setting goals, to monitoring progress,

to evaluating data, to modifying instruction, to evaluating the effectiveness of the modification, to finally determining whether the long-range goal had been met. Less knowledgeable teachers, on the other hand, described the graphs in a mostly random manner.

**Future research directions**

To recap, to this point, I have described three phases of my research program: development of progress measures, effects of implementation of progress measurement, and teachers' use of data. What do I foresee as my future research directions? First, I hope to continue my work on teachers' use of data. This is an area of critical importance for successful implementation of a progress monitoring system. Second, I hope to replicate and extend my work on the development of progress measures for secondary-school students here in the Netherlands. Specifically, I hope to replicate the work on the development of measures in reading and writing, and to explore the development of measures in new areas such as second-language learning and academic-behavior. In this regard, I and my Leiden colleague, Marian Verhallen, have had the good fortune to be able to work with Dr. Kars Veling, Gerieke Til, and the teachers from the Johan de Witt schools in Den Haag on a project in which *leerbiografies* - or learning biographies - are being created for every student in the school. I look forward to our collaborative efforts in the coming years.

**Summary and Implications**

I began my talk with a comparison of the educational systems here and in the United States, and talked about the need to combine the best of both worlds to create a realistic and relentless system of education for students with learning and behavior difficulties. How might such a system be implemented in the schools? In the United States in recent years, such a system has been implemented in school districts throughout the country and has come to be known as *Response to Intervention* or *RTI*. Briefly, RTI involves tiers or levels of interventions in which interventions become increasingly more intensive and specialized (e.g., see L. Fuchs, 2003; D. Fuchs, Mock, Morgan & Young, 2003). The tier or level in which a student is placed, and the decision to move a student from one level to another, is made on the basis of that student's performance and progress.

How could realistic and relentless education be implemented in the Netherlands? With Passend Onderwijs, more and more responsibility is being given to regular schools to create educational programs for more and more students. A data-based decision approach might enable schools to make informed decisions about the success of their programs for students who struggle with learning and behavior. Students likely to experience problems in an academic area - learning English for example - could be identified early and monitored on a regular basis to examine the effects of the regular instructional program on English language learning. If the program is not effective, changes in the program could be made. If the data reveal that repeated, intensive changes do not lead to improvements in performance, consideration could be given to a different program or placement for the student. The performance and the progress of the student would drive the decision-making process, as would the students' response to increasingly intensive interventions.

**Conclusion**

In conclusion, my research journey began 20 years ago in Minnesota, and I hope will continue another 15 to 20 years here in Leiden. The goal of my research program has been to help educators to be both realistic and relentless in their pursuit of effective educational programs for students with learning and behavioral difficulties through the use of data-based decision making. Realistic in the sense that progress data are used to evaluate the appropriateness of educational programs based on growth within those programs rather than predetermined notions about how much the student is likely to learn given a diagnosis or a label. Relentless in the sense that programs and interventions are continuously evaluated and

modified until a successful formula is found for an individual student, even if that formula involves a unique placement or program.

Will we reach the goal of developing programs that are both realistic and relentless? Will we get to where we want to go? It is too soon to say, but to quote the American baseball player, Yogi Berra - the Johan Cruijff of America in terms of creative language use - "If you don't know where you are going, you will wind up somewhere else". We think we know where we are going - but if we are wrong, we will surely end up somewhere else just as interesting!

**Acknowledgements**
In closing, I would like to thank the Leiden University Fund for this special chair. I would also like to thank my University of Minnesota colleagues, Stanley Deno, Kristen McMaster, and Susan Rose, for all of their care, support, and collaborative work during our 20 + years in Minnesota. And I would like to thank my Leiden colleagues for the warm and supportive welcome they have given us here. It is a difficult transition to move from one University to another, especially when that move involves a new country, a new language, and a new system of education. That transition has been made much easier by colleagues such as Paul Vedder, Adrianna Bus, Hannah Swaab, Kees Verduin, Marian Verhallen, Marga de Jong, Marianne van Dijken, Anneke Both-de Vries, Peter van den Bergh, Mitch van Geel, Jill Mos van den Doel, the late Michiel Oortwijn, my chair, Rien van IJzendoorn, Dean Philip Spinhoven, Yvonne van Iperen, and all of the doctoral students and staff from Orthopedagogiek and Pedagogiek, and the staff of FSW.

Of course, I would also like to thank my families for all of the love and support they have given us over the years - both my Dutch family, the Van den Broeks, and my own family, the Espins. I am especially grateful to my Mother and twin sisters, Jan and Julie, for flying over from the United States to share this day with me, and am only sorry that my three brothers and Father could not make the trip. Finally, last but not least, my love and thanks to my husband, Paul, who has always been my source of strength and inspiration, and to our two children Marcus and Marika. Marcus and Marika, you are relentless in your search for truth with questions such as, "What is God's telephone number?" and then as you run to get the telephone book, "What is God's first name, anyway?". Painfully, you are also realistic in your understanding that Mama and Papa do not have all the answers. Simply put: You are the best.

I thank you all for your attention today.
Ik heb gezegd.

**References**


Black, P. & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education, 5,* 7-74.

Black, P. & Wiliam, D. (2005). Lessons from around the world: how policies, politics and cultures constrain and afford assessment practices. *The Curriculum Journal, 16,* 249-261.

Campbell, H., Wallace, T., Lembke, E. & Espin, C.A. (2005). *The effects of parent involvement in progress monitoring using Curriculum-Based Measurement.* Paper presented at the Council for Exceptional Children, Baltimore, MD, April.

Cronbach, L.J. & Meehl, P.E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52,* 281-302.

De Groot, A. D. (1946*). Het denken van den schaker* [Thought and choice in chess]. Amsterdam: North-Holland.

De Groot, A. D. (1965). *Thought and choice in chess.* The Hague: Mouton. (Original work in Dutch published in 1946).

Deno, S.L. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children, 52,* 219-232.

Espin, C.A., Busch, T., Shin, J. & Kruschwitz, R. (2001). Curriculum-based measures in the content areas: Validity of vocabulary-matching measures as indicators of performance in social studies. *Learning Disabilities Research and Practice, 16,* 142-151.

Espin, C.A., De La Paz, S., Scierka, B.J. & Roelofs, L. (2005). Relation between curriculum-based measures in written expression and quality and completeness of expository writing for middle-school students. *Journal of Special Education, 38,* 208-217.

Espin, C.A. & Deno, S.L. (1993). Performance in reading from content area text as an indicator of achievement. *Remedial and Special Education, 14,* 47-59.

Espin, C.A. & Foegen, A. (1996). Validity of three general outcome measures for predicting secondary students' performance on content-area tasks. *Exceptional Children, 62,* 497-514.

Espin, C.A., Scierka, B.J., Skare, S. & Halverson, N. (1999). Criterion-related validity of curriculum-based measures in writing for secondary school students. *Reading & Writing Quarterly, 14,* 5-27.

Espin, C.A., Shin, J. & Busch, T.W. (2005). Curriculum-based measurement in the content areas: Vocabulary-matching as an indicator of social studies learning. *Journal of Learning Disabilities, 38,* 353-363.

Espin, C.A., Skare, S., Shin, J., Deno, S.L., Robinson, S. & Brenner, B. (2000). Identifying indicators of growth in written expression for middle-school students. *Journal of Special Education, 34,* 140-153.

Espin, C.A., Wallace, T., Campbell, H., Lembke, E., Long, J. & Tichá, R. (2008). Curriculum-based Measurement in writing: Predicting the success of high-school students on state standards tests. *Exceptional Children, 74,* 174-193.

Espin, C.A., Wallace, T., Lembke, E., Campbell, H. & Long, J.D. (2010). Creating a progress measurement system in reading for middle-school students: Monitoring progress towards meeting high stakes standards. *Learning Disabilities Research and Practice, 25,* 60-75.

Espin, .C.A., Wallace, T., Long, J., Lembke, E., Campbell, H. & Tichá, R. (2003). *Creating a progress monitoring system: Preparing secondary-school students for success on the Minnesota Basic Skills Tests.* Paper presented at the Pacific Coast Research Conference, La Jolla, CA, February.

Espin, C.A., Wallace, T., Tichá, R., Wayman, M.M., Wiley, H.I. & Long, J. (2006). *Reading general outcome measures: Age and skill-level comparisons.* Paper presented at the Pacific Coast Research Conference, San Diego, CA, February

Espin, C.A., Wayman, M.M., McMaster, K., Deno, S. & De Rooij, M. (2010). *Teachers' understanding of progress monitoring data: Sequential decision making.* Paper presented at the Pacific Coast Research Conference, San Diego, CA, February.

Fuchs, L.S. (2003). Assessing intervention responsiveness:

Conceptual and technical issues. *Learning Disabilities Research and Practice, 18,* 172-186.

Fuchs, D., Mock, D., Morgan, P.L. & Young, C.L. (2003). Responsiveness-to-intervention: Definitions, evidence, and implications for the learning disabilities construct. *Learning Disabilities Research and Practice, 18,* 157-171.

Marston, D. (1989). A curriculum-based measurement approach to assessing academic performance: What is it and why do it. In M. Shinn (ed.), *Curriculum-based measurement: Assessing special children* (pp. 18-78). New York: Guilford.

Messick, S. (1989a). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher, 18,* 5-11.

Messick, S. (1989b). Validity. In R.L. Linn (ed.), *Educational measurement* (3rd ed., pp. 13-103). New York: Macmillan.

Perkins, D.N. (2009). Decision making and its development. In *Education and a civil society: Teaching evidence-based decision making* (pp. 1-28). Cambridge, MA: American Academic of Arts and Sciences.

Simon, H.A. (1990). Invariants of human behavior. *Annual Review of Psychology, 41,* 1-19.

Stecker, P.M., Fuchs, L.S. & Fuchs, D. (2005). Using curriculum-based measurement to improve student achievement: Review of research. *Psychology in the Schools, 42,* 795-819.

Tichá, R., Espin, C.A. & Wayman, M.M. (2009). Reading progress monitoring for secondary-school students: Reliability, validity, and sensitivity to growth of reading aloud and maze selection measures. *Learning Disabilities Research and Practice, 24,* 132-142.

Tillema, H. H. (2009). Assessment for learning to teach; approaisal of pratice teaching lessons by mentors, supervisors and student teachers. *Journal of Teachers Education, 60,* 155-167.

Tillema, H.H. & Smith, K. (2009). Assessment orientation in formative assessment of learning to teach. *Teachers & Teaching: Theory and Practice, 15,* 391-405.

Todd, P.M. (2007). How much information do we need? *European Journal of Operational Research, 177,* 1317-1332.

U.S. Department of Education, Office of Elementary and Secondary Education (2002). *No Child Left Behind: A desktop reference.* Washington, D.C.

Vicente, K.J. & De Groot, A.D. (1990). The memory recall paradigm: Straightening out the historical record. *American Psychologist, 45*(2), 285-287.

Wallace, T., Espin, C.A., Tichá, R., Wayman, M.M., Wiley, H.I. & Long, J. (2005). *The effects of student involvement in progress monitoring using Curriculum-Based Measurement.* Paper presented at the Council for Exceptional Children, Baltimore, MD, April.

Wayman, M.M., Wallace, T., Wiley, H.I., Tichá, R. & Espin, C.A. (2007). Literature synthesis on curriculum-based measurement in reading. *Journal of Special Education, 41,* 85-120.

Weissenburger, J.W. & Espin, C.A. (2005). Curriculum-based measures of writing across grade levels. *Journal of School Psychology, 43,* 153-169.

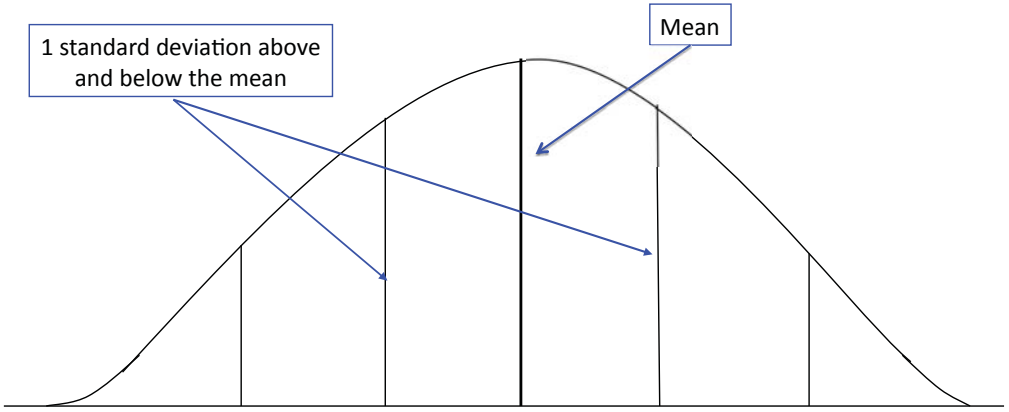


*Figure 1:* Depiction of a normal curve.



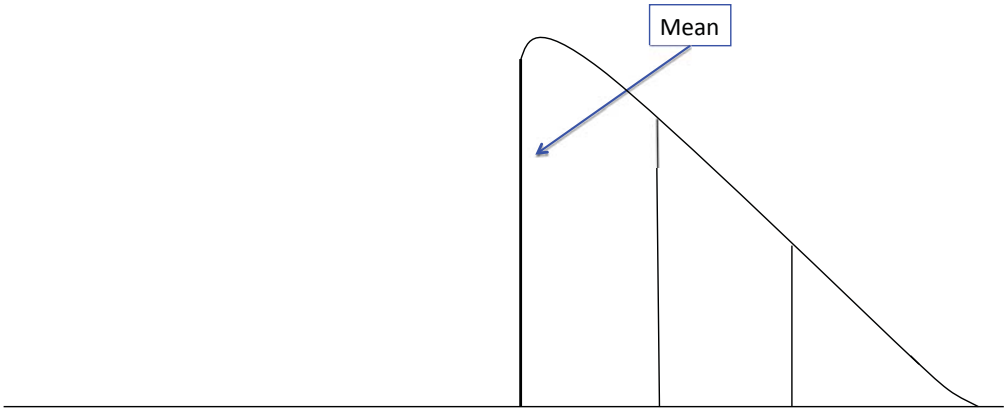


*Figure 2:* Depiction of the U.S. American "version" of a normal curve.

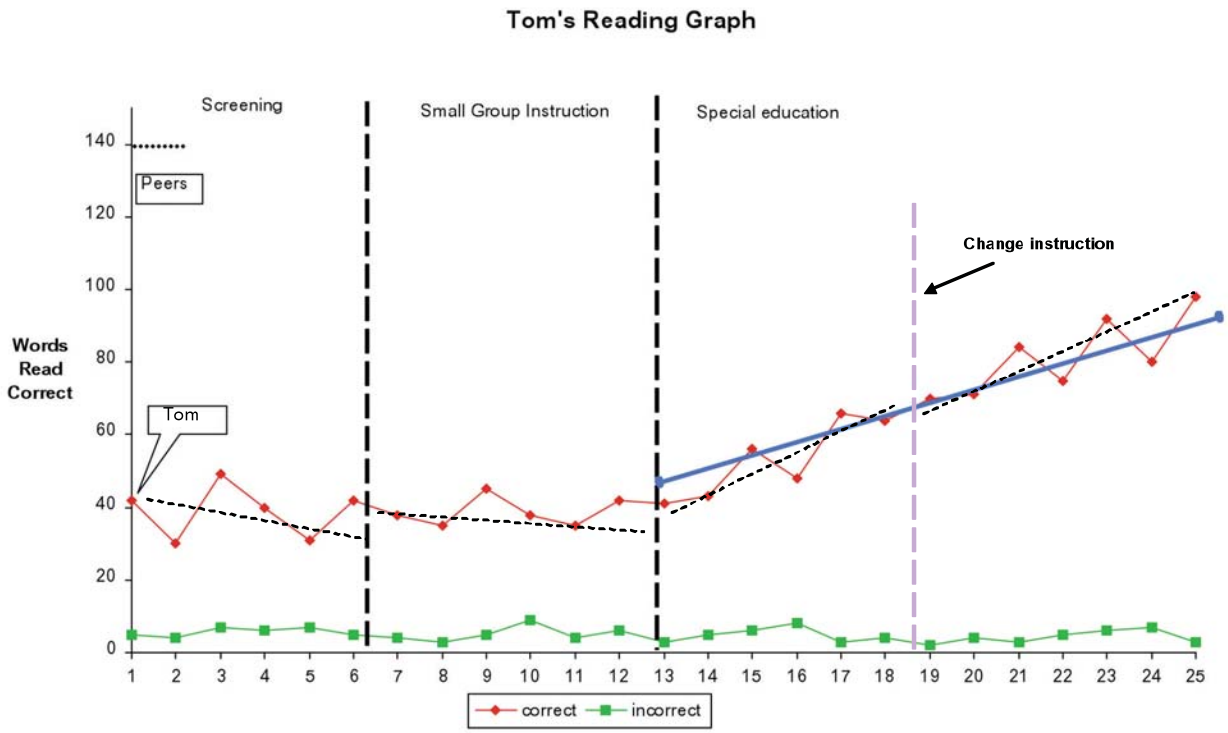


*Figure 3:* Progress graph for Tom.

## Prof.dr. Christine Espin

| | |
|---|---|
| 1990 | Ph. D., Educational Psychology, University of Minnesota |
| 1990-1995 | Assistant Professor, Educational Psychology, University of Minnesota |
| 1995-2002 | Associate Professor, Educational Psychology, University of Minnesota |
| 2002-2008 | Professor, Educational Psychology, University of Minnesota |
| 2008-2009 | Associate Professor, Education and Child Studies, Leiden University |
| 2008-Present | Adjunct Professor, Cognitive Sciences, University of Minnesota |
| 2010-Present | Professor (Bijzondere Hoogleraar), leerstoel, 'Teachers' Assessment of Learning and Behavioral Problems in Adolescents, Leiden University, from Stichting Leids Universiteits Fonds |

Christine Espin's research focuses on the improvement of school performance for secondary-school students with learning and behavior difficulties. Specifically, her research focuses on the development of progress monitoring (Curriculum-based Measurement) procedures that can be used by educators to evaluate the effects of instructional programs on student learning. The first phase of the research focused on the development of measures that would serve as valid and reliable indicators of student performance and progress in reading, writing, and content-area learning at the secondary-school level. The second phase of the research focused on the effects of implementation of progress monitoring procedures on teacher instruction and student performance. The most recent phase of the research focuses on understanding and improving teachers' use of progress monitoring data for decision-making. Dr. Espin began her career as a teacher for secondary-school students with learning and behavioral difficulties.

Universiteit Leiden