

The Challenge of Early Identification of Later Reading Difficulties

by Mads Poulsen

Preventing reading difficulties before they arise is by far preferable to attempting to remediate difficulties after a child has lagged visibly behind. There are two components in preventing reading difficulties: One is to provide strong instruction for everyone. The other is to identify which children should be given additional attention. All children need to learn the same principles and practices of decoding, but some children may require more support and attention, either because it takes them longer to learn the principles with the same amount of effort or because they have difficulty applying the necessary amount of effort by themselves. This article is about the benefits, but more so of the challenges with formalized screening procedures for early identification of children who later develop reading difficulties. The main challenge with early identification is that it tends to be somewhat inaccurate. This is to be expected. Predicting the future is hard. Screening can still be useful, and maybe more so if this limitation is understood and can be taken into account. This article describes the limitations and different approaches to dealing with them.

Early identification of difficulties is beneficial because it allows early support (intervention) and, one hopes, prevention or softening of reading difficulties. “Early” could be defined in many ways. It should be sufficiently early for intervention to be able to prevent a feeling of failure that could extinguish the child’s motivation to keep trying, and before limited reading abilities become a clear impediment to participation in general classroom activities. In this article, early identification will mean *before* actual direct instruction in reading words, but possibly *while* learning foundational skills such as letter knowledge and phoneme awareness.

Predicting Something that Has Not Happened is Difficult

The challenge of early identification before instruction is that it involves predicting something—reading development and difficulty—that has not happened yet. Once instruction is underway, children’s present reading ability becomes a strong predictor of their future reading ability and difficulty: If children struggle to read after a certain amount of instruction, then the risk is high that they will continue to struggle. But before instruction in reading words, reading ability is often not a very good indicator of later reading ability. Children who can read without instruction are probably in the clear. But it does not tell much about a child that he or she cannot read before being taught how!

There are other and better early indicators of later reading performance. For example, there is a statistical tendency for children with poor letter knowledge and phoneme awareness to have difficulties years later with reading. However, it is a tendency, not a certainty. Some children with poor letter knowledge catch up just fine after some time. The consequence is that an early screening will flag some children as being at risk, who

will turn out to be fine. These cases are sometimes referred to as *false positives* because they are falsely flagged as being positive for possible reading difficulty. On the other hand, some children who actually do develop reading difficulties will be overlooked. These are sometimes referred to as *false negatives*. In other words, early identification of later reading difficulties is not going to be 100% precise. Most of us accept solutions that do not work perfectly all of the time. Knowing how well a solution can be expected to work allows us to take reasonable precautions.

Most of us accept solutions that do not work perfectly all of the time. Knowing how well a solution can be expected to work allows us to take reasonable precautions.

How to Evaluate Identification Accuracy

One way of evaluating how well an early screening procedure works is to use the procedure with a group of children before or in the beginning of reading instruction, and then revisit the children after a few years to see who developed reading difficulties. It is then possible to compare who the screening procedure predicted to have difficulties with who actually developed difficulties. This comparison can be made in many ways. The simplest statistic is unfortunately not very good: It is tempting to simply calculate the percentage of children who were correctly classified by the screening as having reading difficulties or not. For example, an early screening might classify 92% of the children correctly. This may sound impressive, but if 8% of the children ended up with reading difficulties, the test could accomplish a 92% classification accuracy simply by predicting that nobody would develop reading difficulties! No legitimate screening procedure would do this, but it shows that this simple statistic can be very misleading. Instead, there exists a number of different and complementary statistics, each highlighting different aspects of identification accuracy.

Since the goal of screening is to identify children who need special instructional attention, it is useful to know how well a screening tool does this. The *sensitivity* is the percentage of children who are correctly flagged for being at risk (*true positives*) out of all the students who will experience difficulties. Since the aim is to identify children in need of attention, this percentage should preferably be as high as possible.

But the screening should also avoid raising unnecessary concern associated with flagging children as being at risk when

Continued on page 12

in reality they are going to catch up just fine. The *false positive rate* (also known as *1-specificity*) is the percentage of children who will not develop difficulties, but who are incorrectly identified as being at risk. The aim is to keep the false positive rate low. It is important to keep in mind that at the point in time of early screening, there is no way of differentiating between true and false positives. Besides raising unnecessary concern, false positives may tie up resources that could have been used more productively elsewhere. For example, the more children a teacher has to direct special attention to, the less intense that attention is probably going to be.

Trade-off Between Finding Those in Need and Raising Unnecessary Concern

For a given screening tool, the sensitivity and the false positive rate are determined by the cut-off test result that determines whether a child is flagged as at risk or not. The cut-off could, for example, be a certain number of letters on a letter knowledge test. People, not nature, decide cut-offs. The decision is made difficult by the fact that there is a trade-off between the sensitivity and the false positive rate. If the cut-off is set such that many children will be flagged as at risk (e.g., they should know at least 24 letters by school entry, otherwise they are flagged), then many of the students who will eventually have reading difficulties will be flagged appropriately: The sensitivity will be high, which is the goal. However, the flip side is that many students who will turn out not to have reading difficulties will be inappropriately flagged, too. The false positive rate will unfortunately also be high. For example, in one study where we followed a group of Danish students from kindergarten to grade 2, we found that an early screening procedure that set the cut-off to produce a sensitivity of 80% resulted in a false positive rate of 29% (Poulsen, Nielsen, Juul, & Elbro, 2017). In other words, if the ambition was to early identify 80% of those who would require special attention in grade 2, that would entail also flagging for special attention about a third of the children who would do well if they received no special attention.

For a given screening procedure there is a tough choice between finding as many as possible of the children who need the attention and ending up with a manageable group size where not too many of the children might better spend their time elsewhere.

If we instead chose a less ambitious cut-off corresponding to a 60% sensitivity, then the false positive rate would improve; it would go down to 14%. But that would be at the cost of overlooking more students who likely would have benefited from some special teaching. These results are not very different from

other studies of attempted identification prior to formal reading instruction (Sittner Bridges & Catts, 2011; Catts, Petscher, Schatschneider, Sittner Bridge, & Mendoza, 2009; Gellert & Elbro, 2017; Johnson, Jenkins, Petscher, & Catts, 2009).

Thus, for a given screening procedure there is a tough choice between, on the one hand, finding as many as possible of the children who need the attention, and, on the other hand, ending up with a manageable group size where not too many of the children might better spend their time elsewhere.

It is natural to have high ambitions for the sensitivity. After all, the goal is to find those who need attention. But finding the students in need is only the first step. They also need to be given the attention. And if too many are flagged as being at risk, then it may be difficult to give the necessary amount of attention to each child, unless one's education budget is infinite.

From a day-to-day practical standpoint, it might be useful to keep an eye on another statistic. The *total positive rate*, that is the overall percentage of children who are flagged as being at risk, disregarding whether they end up with difficulties or not. This is the number of children who the school has to provide with special instructional attention. On a fixed budget, the total positive rate may thus dictate what kind of special attention is feasible. For example, it may not be possible to offer many hours of one-on-one instruction for, say, one third of the students in every classroom.

The total positive rate is heavily influenced by the false positive rate because most children do not end up having reading difficulties. It can be calculated from the prevalence, sensitivity, and false positive rates. In the study noted above (Poulsen et al., 2017), the prevalence of reading difficulty was 15%. If we aimed at 80% sensitivity, we could expect a false positive rate of 29%. Thus, out of 100 students, about 25 would be false positives $((100-15) \times 0.29)$, and 12 (15×0.80) would be true positives in actual need of attention, for a total of 37 students flagged for attention. If the aim was a more modest 60% sensitivity, then the total positive rate would be a more manageable 21 students flagged for attention $((100-15) \times 0.14) + (15 \times 0.60)$.

Improving Screening Accuracy

Some screening procedures are better at predicting future reading difficulties than others. They have more favorable trade-offs between sensitivity and false positive rates, meaning that it will be possible to set more ambitious goals for finding those in need, while keeping the false and total positive rates at manageable levels.

There is important research that has shown how early screening accuracy can be improved. Accurate screening procedures typically use tests of multiple foundational skills. Lately, tests that measure how well children learn principles of reading from brief instruction during a testing session have shown very promising results (Gellert & Elbro, 2017). But still, early screening cannot yet be assumed to be so precise that the practical problem of the trade-off disappears.

Addressing the Trade-Off: Two Scenarios

Again, the challenge of early, relatively inaccurate screening leaves a tough choice between two goals: identifying as many children in need as possible (high sensitivity) versus not raising unnecessary concern and having enough resources to actually help the children who are identified as being in need (low false positive and total positive rates). By acknowledging this trade-off, it may be possible to anticipate and meliorate the problems associated with different approaches to the trade-off.

One approach to the trade-off is to prioritize the identification of as many of those who will develop reading difficulties as possible. In other words, failing to identify a child with learning needs may be thought to be a worse outcome than the problem of falsely identifying a child who really did not need extra assistance. This strategy involves setting cut-offs with the aim of achieving high sensitivity, despite the associated high false and total positive rates. Say a third of an average classroom could be flagged for attention, as in the above example that aimed at finding 80% of the students who would develop reading difficulties. On a fixed budget, this probably constrains the intensity of the intervention. To accommodate this situation, the intervention could consist of dividing the classroom into two or three groups who work on different tasks that are suitable for different reading levels (see Connor et al., 2013 for a similar approach with first- to third-grade students). Communication with students and parents should be mindful of the fact that many of the students who are flagged for attention will not develop reading difficulties.

One approach to the trade-off is to prioritize the identification of as many of those who will develop reading difficulties as possible. Another approach is to prioritize providing relatively intensive interventions for the few students who are most likely to develop reading difficulties.

Another approach is to prioritize providing relatively intensive interventions for the few students who are most likely to develop reading difficulties. This strategy involves setting the screening cut-off value at a level where the total positive rate is low enough to allow the desired intensity of intervention within the available resources. Such an intervention can be organized in many ways, from simply letting a proficient classroom teacher focus more time on these few students and maybe including the parents in the efforts, to providing pull-out individual or small group instruction with specialized teachers. In the above example, when a cut-off was set to produce a true positive rate of 60%, the total positive rate was 21% or about five students in a classroom of 25. Possibly not low enough for intensive individual instruction, but small group instruction might be manageable in such a case. Of course, the downside to this approach is that many of the children who will develop difficulties will not be flagged. These may be children who

did not do too badly on the letter knowledge and phoneme awareness tasks, but who for some reason later struggle with learning the more complex task of reading actual words of increasing difficulty. In any case, this approach can be expected to overlook a substantial number of the children who will eventually develop difficulties. To offset this, it would be necessary to keep a close eye on how *all* the children progress.

After instruction begins, the students' performance on actual reading tests becomes a better predictor of later reading difficulty (Catts et al., 2009; Compton et al., 2010; Poulsen et al. 2017), allowing more precise identification. In the Danish example (Poulsen et al., 2017), much better identification accuracy was achieved with relatively simple word-list reading measures in the January of grade 1: for example, 80% sensitivity with an 8% false positive rate. In another study, Compton and colleagues (2010) demonstrated even better identification accuracy of reading difficulties in the end of grade 2 with measures that were administered in the fall of grade 1: about 90% sensitivity with a 10% false positive rate. However, it should be noted that this impressive result came at the cost of a rather intensive testing procedure involving weekly progress monitoring with word-list reading measures for five weeks. At some point, the benefit of small improvements in screening accuracy should be weighed against the cost of disrupting ordinary instruction and collecting and managing the data—especially considering that the identification accuracy with simple tools increases automatically as instruction progresses, and it becomes easier to notice which students are not responding to ordinary instruction or specialized intervention.

Screening does not have to be a one-off affair. Early, inaccurate screening can, and probably should, be supplemented with follow-up screening or monitoring at suitable intervals. The intervals could be long enough to allow some of the slow learners to progress meaningfully, but short enough to allow teachers to catch students who are not responding or find new challenges for students who are responding.

Screening does not have to be a one-off affair. Early, inaccurate screening can, and probably should, be supplemented with follow-up screening or monitoring at suitable intervals.

The two above approaches to *early* identification can be seen as early starts that either prioritize giving some instruction to the many who need it (and many who probably do not), or prioritizing giving more intensive instruction to a few who can be expected to need it the most. Which approach to take will depend on many factors in a school system. But knowing the limitations of early screening allows finding a solution that fits the students and resources of particular schools.

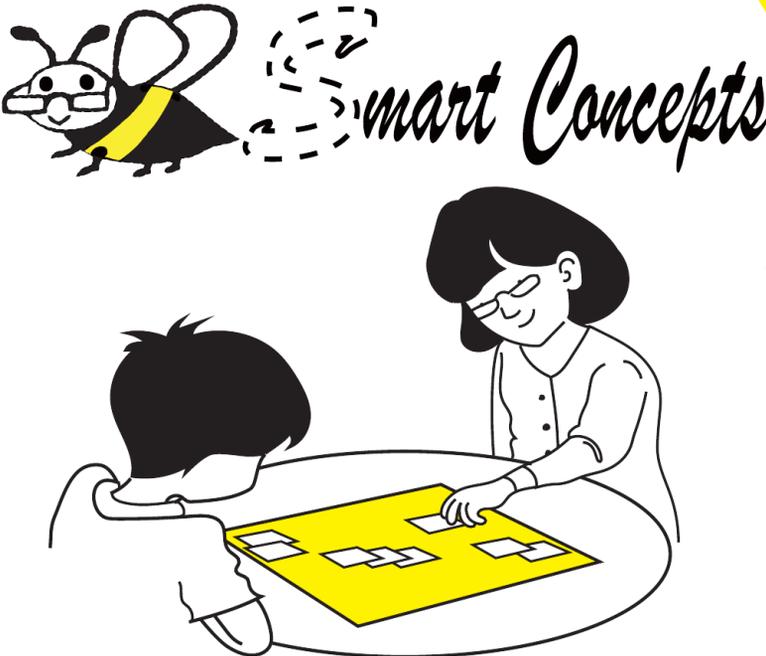
Continued on page 14

References

- Catts, H. W., Petscher, Y., Schatschneider, C., Sittner Bridges, M., & Mendoza, K. (2009). Floor effects associated with universal screening and their impact on the early identification of reading disabilities. *Journal of Learning Disabilities, 42*, 163–176. <http://dx.doi.org/10.1177/0022219408326219>
- Compton, D. L., Fuchs, D., Fuchs, L. S., Bouton, B., Gilbert, J. K., Barquero, L. A., Cho, E., & Crouch, R. C. (2010). Selecting at-risk first-grade readers for early intervention: Eliminating false positives and exploring the promise of a two-stage gated screening process. *Journal of Educational Psychology, 102*, 327–340. <http://dx.doi.org/10.1037/a0018448>
- Connor, C. M., Morrison, F. J., Fishman, B., Crow, E. C., Al Otaiba, S., & Schatschneider, C. (2013). A longitudinal cluster-randomized controlled study on the accumulating effects of individual literacy instruction on students' reading from first through third grade. *Psychological Science, 24*, 1408–1419. <http://dx.doi.org/10.1177/0956797612472204>
- Gellert, A. S., & Elbro, C. (2017). Try a little bit of teaching: A dynamic assessment of word decoding as a kindergarten predictor of word reading difficulties at the end of grade 1. *Scientific Studies of Reading, 21*, 277–291. <http://dx.doi.org/10.1080/10888438.2017.1287187>
- Johnson, E. S., Jenkins, J. R., Petscher, Y., & Catts, H. W. (2009). How can we improve the accuracy of screening instruments? *Learning Disabilities Research & Practice, 24*, 174–185. <http://dx.doi.org/10.1111/j.1540-5826.2009.00291.x>
- Poulsen, M., Nielsen, A.-M. V., Juul, H., & Elbro, C. (2017). Early identification of reading difficulties: A screening strategy that adjusts the sensitivity to the level of prediction accuracy. *Dyslexia, 23*, 251–267. <http://dx.doi.org/10.1002/dys.1560>
- Sittner Bridges, M., & Catts, H. W. (2011). The use of a dynamic screening of phonological awareness to predict risk for reading disabilities in kindergarten children. *Journal of Learning Disabilities, 44*, 330–338. <http://dx.doi.org/10.1177/0022219411407863>

Mads Poulsen, Ph.D., is an associate professor at the University of Copenhagen, Denmark, and is associated with the Centre for Reading Research. His research interests include identification of reading difficulties and the linguistic underpinnings of reading development. You can write to Mads Poulsen at m.poulsen@hum.ku.dk

Advertisement



Bee Smart Concepts

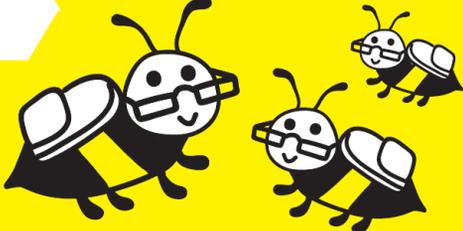
Bee Smart Concepts is the creator of illustrated children's games that support the Orton-Gillingham approach.

Our program promotes **fluency** and **phonological awareness** through engaging, repetitive card and board games.

Games offered by Bee Smart:

- Snap Games
- Sound Fishing Card Games
- Crazy Syllables Card Games
- Board Games

For more information about our games please visit our website @ www.beesmartconcepts.com



Advertisement

IDA Fact Sheets Now Available in Spanish

A series of fact sheets from IDA aims to help students and parents understand and apply for accommodations at all levels of education. Our K-12 fact sheet, Accommodations for Students with Dyslexia, is also an excellent overview of accommodations for all ages. For a guide to accommodations at colleges and universities, see *Applying for Accommodations on College Entrance Tests* and *Applying for Accommodations on Graduate School Entrance Tests*. Visit <https://DyslexialDA.org/fact-sheets/> to download and share.