

## **FAQ #1: How & Why did you select your instructional standards (i.e., frustration, instructional, & mastery criteria)?**

### **Background & Key Terms**

To begin answering your questions, let's establish two crucial terms: **Testing** and **Evaluation**.

1. **Testing** is the process of collecting student **responses** (or performance) when they are presented with academic tasks or stimuli. It is the act of measurement.
2. **Evaluation** is the **judgment** we make using that testing data to drive educational decisions (such as intervention placement or instructional change).

For evaluation to be meaningful, the collected responses must be compared objectively. This is where **Standards** come in. Standards define the expected **levels** or **rates** of performance we want students to exhibit. We use these standards to guide our judgments about the degree of the gap, or discrepancy, between a student's actual performance and the expected performance.

There are two types of standards commonly used in education: **Norm-Referenced** and **Criterion-Referenced**. Norm-Referenced standards compare a particular student's performance against that of a peer group, essentially providing a rank order of data. In contrast, Criterion-Referenced standards measure a student's performance relative to a fixed, expected skill level on a specific task. Both methods are strengthened through **standardization**, which validates the data to ensure they reliably meet their intended purpose, such as instructional placement or screening. For the purpose of this Q&A series, when we discuss standards, we will primarily be referring to the **Criterion-Referenced** standards used for instructional placement and evaluation.

One **Criterion-Referenced Standards** used for instructional placement defines three distinct levels of performance: **Frustration**, **Instructional**, and **Mastery** (Coddling et al., 2017). The **Frustration Range** indicates the student has low levels of skill performance and will need intensive support; further assessment should be conducted to verify if critical prerequisite skills are missing, examining both the student's accuracy and fluency. The **Instructional Range** is where students exhibit the target skill with a functional level of accuracy and fluency; these students are ideally positioned to respond well to focused interventions, particularly those involving high rates of retrieval practice and performance feedback, which help them move toward mastery. Finally, **Mastery** is achieved when the student's fluent responding is high enough that it ensures the target skill will likely maintain over time (VanDerHeyden & Burns, 2009) and allows them to learn future, related skills with increased efficiency (Schutte, 2015).

### **Published Standards**

Frustration, Instructional, and Mastery criteria vary depending on who you ask and these also vary across grades and skills. In the book, *Effective Math Interventions* (Coddling et al., 2017), the authors highlighted three sets of commonly cited standards.

**TABLE 3.2. Skill-by-Treatment Interaction: Combining CBA and the Instructional Hierarchy for Treatment Selection**

| Performance level | CBA criterion            |                       |                                     | Stage of instructional hierarchy              |
|-------------------|--------------------------|-----------------------|-------------------------------------|-----------------------------------------------|
|                   | Deno & Mirkin (1977)     | Howell & Nolet (1999) | Burns, VanDerHeyden, & Jiban (2006) |                                               |
| Frustration       | Grades 1–3<br><10 DCPM   | <20 DCPM              | Grades 2–3<br><14 DCPM              | Acquisition                                   |
|                   | Grades 4+<br><20 DCPM    |                       | Grades 4–5<br><24 DCPM              |                                               |
| Instructional     | Grades 1–3<br>10–19 DCPM | 20–30 DCPM            | Grades 2–3<br>14–31 DCPM            | Fluency                                       |
|                   | Grades 4+<br>20–39 DCPM  |                       | Grades 4–5<br>24–49 DCPM            |                                               |
| Mastery           | Grades 1–3<br>20+ DCPM   | 40+ DCPM              | Grades 2–3<br>32+ DCPM              | Maintenance;<br>generalization;<br>adaptation |
|                   | Grades 4+<br>40+ DCPM    |                       | Grades 4–5<br>50+ DCPM              |                                               |

Deno and Mirkin (1977) provided some of the initial published standards. They separated criteria by performance and grade levels (see above). For grades 1-3 mastery was set at 20+ DCPM and for grades 4-6 the ranges doubled and mastery was set at 40+ DCPM.

Howell & Nolet (1999) kept it simple and just used the Deno & Mirkin (1977) criteria for grades 4-6 and generalized it for all basic fact skills and grades (I made a typo in the book table for the Howell & Nolet instructional range...it is 20-39 DCPM, not 20-30 DCPM).

Burns et al (2006) conducted a robust analysis from test-retest data and intervention growth data to derive instructional ranges. Instructional levels were 14-31 DCPM for grades 2-3 and 24-49 DCPM for grades 4-5. VanDerHeyden & Burns (2009) later published another study where they looked at DCPM scores to predict retention with grades 2 and 3 retaining skill at approximately 24-34 DCPM and grades 4-5 retaining skills at approximately 60 DCPM (depending on skill).

### **Teacher Issues with Standards**

As we moved toward selecting our standards, we looked to previously cited research but also considered teacher feedback. When we started the M.I.N.D. we used DCPM criteria across skills and operations. In response to this, teachers provided us with feedback, and three common themes were identified: 1) Difficulty Scoring, 2) Partial Credit for Incorrect Answers, & 3) Criteria was too high.

**Difficulty Scoring.** The use of DCPM, ubiquitous with Curriculum Based Measurement (CBM), got off to a rocky start. I received many emails from teachers who found it was difficult and/or confusing to score. To address this, we made an administration & scoring guide, in hindsight, this was not the fix they were looking for as we continued to receive less than stellar feedback on the response effort that went into weekly scoring (even with the administration & scoring guide).

**Partial Credit for Incorrect Answers.** A recurring theme from teachers was giving students partial credit for a wrong answer. Scoring procedures from past researchers (Deno & Mirkin, 1977) were used to drive these decisions. For those who are curious, DCPM was used to increase the variance, or spread, of scores to support reliability. This is important when scoring multidigit math probes to establish norm referenced scores in manageable amounts of time.

**Criterion is Too High.** Probably the most common question was whether 40 DCPM was an obtainable goal for all students. Teachers were especially perturbed when students were required to do 40 DCPM in subtraction/division since kids had to complete more problems in those operations than with addition/multiplication to reach the criterion. To address this, we lowered criterion for subtraction/division to 30 DCPM and kept the 40 DCPM for addition/multiplication. Unfortunately, now our performance criteria lacked uniformity, complicating interpretation.

### **Selecting Standards for Facts on Fire**

We created Facts on Fire for two reasons. **First**, as School Psychologists, whenever we collected school level norms fact fluency deficits were a Tier 1 issue. Almost all kids failed to hit mastery criteria. This dysfluent performance is widespread (Stickney et al., 2012) and is due to curricula that fail to provide structured practice opportunities for students to perform at mastery levels (Witzel & Riccomini, 2007). **Second**, teachers needed a systematic set of empirically supported resources to promote fluency building in basic math skills that was accessible. Facts on Fire needed to be trainable, easy to roll out, sustainable, and effective. One important facet of the program was to carefully select criteria to guide decisions about when students reached “mastery” that would be clear and interpretable for teachers. To achieve these goals, we selected mastery standards with three principles in mind: 1) Parsimony, 2) Uniformity, & 3) Validity.

Since a primary goal of the program was teacher accessibility, we first focused on adhering to the principles of **parsimony & uniformity**. If we could keep performance standards simple and consistent, teachers could focus efforts on other aspects of program implementation and management. While this is important, we also had to select criteria that were **valid** for the purposes of the program. Specifically, we needed mastery criteria ambitious enough that, once achieved, students would maintain and generalize target skills.

### **Recalibrating Mastery Criteria for Facts on Fire**

For the **M.I.N.D. 2.0 - Facts on Fire**, we kept our goals of parsimony, uniformity, and validity and selected a single criterion for grade 1 (20+ PPM) and grades 2-5 (30+ PPM) that covers all basic fact operations. For students to meet this they need to correctly answer a basic fact every 3 seconds (grade 1) or 2 seconds (grades 2-5). Our decision for switching from DCPM to PPM criteria was twofold: 1) to provide a consistent target level (30 PPM) and 2) to simplify scoring. These were in direct response to teacher feedback and addressed previous concerns.

**Difficulty Scoring & Partial Credit.** The new criteria require teachers and/or students to just count the number of problems correctly completed. No more digits for basic facts, answers are either right or wrong, and no more partial credit. To reduce teacher time scoring probes, we are having students score their daily performance to guide progress monitoring and goal setting procedures. Teachers are no longer asked to score and record student performance each week.

**Criterion is Too High.** This newly selected 20+ (grade 1) and 30+ (grades 2-5) PPM mastery criteria align with previous research based estimates (see Burns et al., 2007; VanDerHeyden & Burns, 2009) discussed earlier in the Q&A and are consistent with recommendations that kids answer their basic facts within 3 seconds per problem in grades K-1 and within 2 seconds grades 2-5. For a detailed breakdown of how PPM criteria convert to DCPM see the table below.

| Skill          | Grades | Facts on Fire PPM Criterion | Facts on Fire PPM to DCPM | Other Criteria |
|----------------|--------|-----------------------------|---------------------------|----------------|
| Sums to 6      | K-1    | 20 PPM                      | 20 DCPM                   | 20+ DCPM*      |
| Sums to 10     | K-1    | 20 PPM                      | 26 DCPM                   | 20+ DCPM*      |
| Sums to 10     | 2-3    | 30 PPM                      | 38 DCPM                   | 34+ DCPM**     |
| Sums to 18     | 2-3    | 30 PPM                      | 50 DCPM                   | 34+ DCPM**     |
| Subtraction 18 | 1-3    | 30 PPM                      | 30 DCPM                   | 32+ DCPM***    |
| Mult to 81     | 3-5    | 30 PPM                      | 57 DCPM                   | 50+ DCPM***    |
| Division 81    | 3-5    | 30 PPM                      | 30 DCPM                   | 50+ DCPM***    |

\* Deno & Mirkin (1977)

\*\* Burns et al. (2007)

\*\*\* VanDerHeyden & Burns (2009)

### **Future Directions: Using Data to Inform & Improve Practice**

We hope our FAQ about how and why we selected performance criteria has been clarified. It is our goal to use empirically informed practices while being sensitive to the challenges that teachers encounter in their classrooms. We will end the FAQ by highlighting some future plans to further investigate this topic in hopes of providing data to inform and improve the Facts on Fire program moving forward.

**Should Students Score their Daily Facts on Fire Timing?** Teachers have expressed concerns that students may be unknowingly counting wrong answers as correct following their daily Facts on Fire retrieval practice session. To address this, we are currently analyzing data that will provide data on their accuracy of responding and how accurately they score and record their PPM scores on the 1-minute Facts on Fire activity. In the meantime, we remind teachers to circulate throughout the room during the intervention to spot-check student accuracy for both. If you have specific concerns about a student's scoring, it would be prudent to score the probes yourself to evaluate accuracy. We are hopeful that our current recommendation will be successful for most students and save teachers valuable time for other instructional activities.

**Will Instructional Levels be Predictive of Student Learning?** We are collecting data on students' response to the Facts on Fire intervention program and we will be looking at student growth patterns across skills. Student growth trajectories will be examined to determine if predictable patterns can be established (e.g., does initial fluency level predict levels and/or rates of improvement). We will also be looking into finding out if a lack of response can quickly be detected to inform instructional recommendations (i.e., increase dose). If predictable patterns emerge, these could be used within a Multi-Tiered System of Support to guide instructional decision making with an emphasis on prevention. It is hoped that these data will help us select and interpret performance criteria to quickly and effectively increase student math achievement.

## References

- Burns, M. K., VanDerHeyden, A. M., & Jiban, C. L. (2006). Assessing the instructional level for mathematics: A comparison of methods. *School Psychology Review*, 35(3), 201-418.
- Codding, R. S., Volpe, R. J., & Poncy, B. C. (2017). *Effective math interventions: A guide to improving whole number knowledge*. Guildford Press.
- Deno, S. & Mirkin P. (1977). *Data-based program modification: A manual*. University of Minnesota.
- Howell, K. W., & Nolet, . (1999). *Curriculum-based evaluation: Teaching and decision making* (3<sup>rd</sup> Ed.).
- Schutte, G. (2015). *Evaluating the role of programming common stimuli in declarative and procedural skills to promote generalization of basic math skills*. Oklahoma State University.
- Stickney, E. M., Sharp, L. D., & Kenyon, A. S. (2012). Technology-enhanced assessment of math fact automaticity: Patterns of performance for low and typically achieving students. *Assessment for Effective Intervention*, 37(2), 84-94.
- VanDerHeyden, A. M. & Burns, M. K. (2009). Performance indicators in math: Implications for brief experimental analysis of academic performance. *Journal of Behavioral Education*, 18: 71-91. DOI 10.1007/s10864-009-9081-x