GROWTH

GUIDE

Using MAP Growth for gifted and talented service placement decisions

Table of contents

Executive summary

School districts should strive to obtain the maximum possible benefit out of a piece of assessment data. To that end, these guidelines were designed to help guide districts in how, when, and why they can use MAP® Growth™ data to inform placement decisions in advanced learning opportunities, such as gifted and talented programs.

MAP Growth can serve as a universal screener (phase one) to determine which students should move forward to the actual eligibility determination phase of program placement decisions. It can also be used as one of the data points on which eligibility decisions are made (phase two). It can make for an ideal universal screener because 1) it has the ability to yield highly reliable data, 2) it is strongly correlated with many phase-two eligibility criteria, and 3) using it as such would require no additional time or resources since MAP partners already have MAP Growth scores for all students.

The most important criteria for when or whether to use MAP Growth as a phase-one screener is whether MAP Growth scores are strongly correlated with the actual placement criteria. If the two phases measure different things and, therefore, the scores on MAP Growth are weakly correlated with scores on the actual placement criteria, then MAP Growth will not work well as a screener.

The most important criteria for when or whether to use MAP Growth as one component of the placement criteria is whether the content measured on MAP Growth are necessary for success in the service. Participation in a service should not be conditional on high MAP Growth scores unless such scores are predictive of success in the service. For obvious reasons, services where high MAP Growth scores make sense as a part of the placement criteria are advanced or accelerated courses in math or that require well-above average reading proficiency.

In the following pages we expand on these guidelines, provide several examples from real school districts, and also address some of the most-frequently asked questions about MAP Growth and advanced learning placement decisions.

Introduction

This document offers guidance for where and how NWEA® MAP Growth can help educators decide which students are ready for or would benefit from advanced learning opportunities, such as gifted and talented programs. Drawing on established research, we outline general best practices for partners to consider followed by specific examples from NWEA partners. Despite its approval as an identification tool in a number of states, it is important to note that MAP Growth was not specifically designed or validated for this purpose. As such, in following the Standards for Educational and Psychological Testing (Standard 12.10), MAP Growth scores are best used alongside multiple datapoints to make service placement decisions, and decisions should always be made in consultation with state regulations. By the end of this document, readers should be able to do the following:

1. Understand what MAP Growth is intended to measure
2. Know the essential components of effective and defensible service placement criteria
3. Understand how, when, and why MAP Growth can be used as a phase-one universal screener
4. Understand how, when, and why MAP Growth can be included as one of several data points when making service eligibility decisions

What are advanced learning opportunities?

“Advanced learning opportunities” are any program, service, instructional intervention, or course that exposes a student to content most often meant for older students or those in higher grades. For example, Algebra I is typically taken in ninth grade, but an increasing number of students are taking it in earlier grades. Similarly, gifted and talented programs are advanced learning opportunities if the goal is to provide exposure to advanced or accelerated academic content.

What is MAP Growth and what does it measure?

MAP Growth is an interim adaptive test used in K–12 to measures a student’s achievement and growth in math, reading, language usage, and science. Because of its computer-adaptive design and vertical scaling, MAP Growth can precisely measure student achievement regardless of grade level.

Effective placement criteria should catch all the students who would benefit from advanced opportunities, avoid considering factors that are irrelevant to readiness or success, and do so all while balancing cost with sensitivity (Peters et al., 2023; Standard 12.13). The guidelines offered here are intended to show how MAP Growth can improve placement decisions along these four "CASA" criteria. Placement criteria should:

- **Cost:** use no more resources (time and/or money) than necessary
- **Alignment:** be aligned in content, domain, and level with the services into which students will be placed
- **Sensitivity:** correctly catch students who would benefit from the service
- **Access:** remove any implicit or explicit barriers that are unrelated to need for or success in the service

Where can MAP Growth play a role?

The Standards for Educational and Psychological Testing and the National Association for Gifted Children Pre-K–Grade 12 Programming Standards emphasize that any program placement decision should rely on multiple data points. The challenge in implementing this best practice is that administering multiple assessments to all students is time consuming and expensive. As a result, identification systems often proceed in two phases to determine which students are ready for or would benefit from an advanced learning opportunity. The first phase is meant to decide which students should go through the eligibility determination phase, which then determines who is eligible for the advanced learning opportunity. For example, students might need to take a test and receive a letter of recommendation to enroll in an accelerated course, but they only take that test and request the letter if they are first referred by a teacher or parent. The letter of recommendation and test score determines if the student is eligible for the course, but it's the teacher or parent referral that initiated the process in the first place.

Rather than testing and collecting letters of recommendation on all students, this kind of two-phase approach to service eligibility reduces costs by limiting the number of students considered for placement in phase two. Fewer tests need to be administered and fewer teachers need to write letters of recommendation. But this can also cause students to be missed since they never had access to phase two. In the example from above, the referral may hold some students back who would have done well on the test and received excellent recommendations. This is particularly concerning given students of color and those from low-income families are less likely to be referred and, therefore, most likely to benefit from a universal screening system that does not rely on such referrals (see Card & Giuliano, 2016; Hyman, 2016; McBee et al., 2016).

Using MAP Growth as a phase-one universal screener

One option to mitigate the downsides of subjective, nonuniversal referrals is to replace them with a universal screener. Figure 1 depicts this process. In the first phase, all students are screened, and those who meet some predetermined criteria or cut score are given further consideration at phase two. In phase two, multiple data points are collected to make decisions about program placement. When a universally administered assessment is used as a phase-one screener, schools can avoid missing students simply because they were never referred for consideration. In short, it makes sure every student has access to the eligibility process, even if not all students go through phase two. Schools that use MAP Growth typically administer the assessment to all students in a grade, which makes it an ideal candidate for use as a phase-one universal screener to determine which students should be considered further.

For schools already using MAP Growth, incorporating it into a two-phase identification system as a universal screener can lead to cost savings given it alleviates the need to administer a separate universal screener. It can also save instructional time since students would not need to sit a second test. By using MAP Growth at phase one, schools can obtain the cost benefits of a two-phase identification system with the sensitivity and access benefits of a single-phase system—they can make sure they are missing as few advanced learners as possible while devoting as few additional resources to testing and identification as possible. By allocating fewer resources to identification assessments in second grade, for example, a school can reallocate those resources to services, or they can expand identification to another grade level, thereby increasing sensitivity and improving access even more.

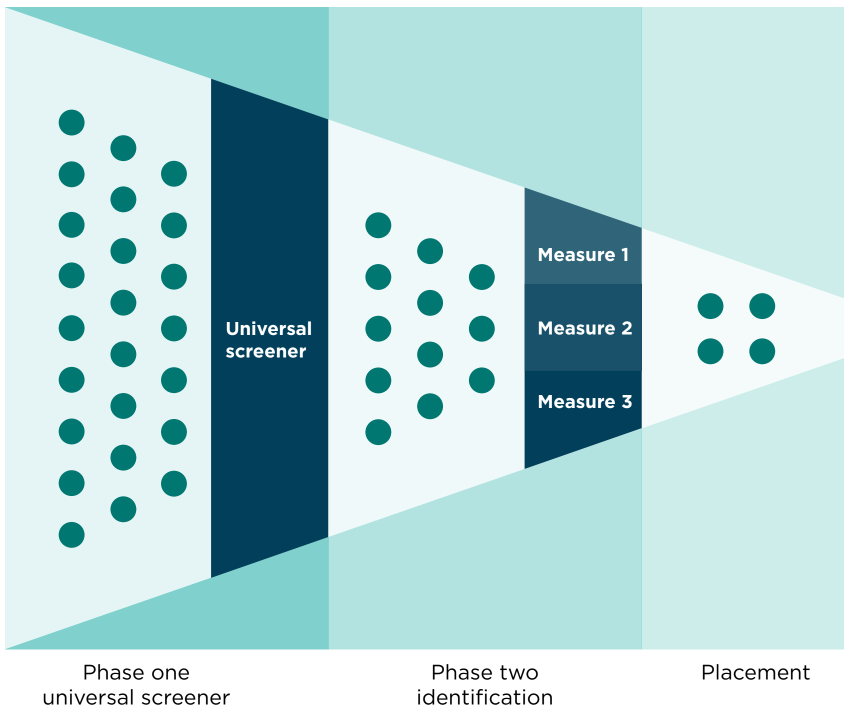


Figure 1. Two-phase identification system with universal screener

Incorporating MAP Growth in a multiple measures approach in phase two

MAP Growth can also play a role at phase two—in making the actual placement or eligibility decisions for certain programs or services. As noted above, program placement decisions should always rely on multiple data points. The most common data points used when making gifted and talented placement decisions include measures of academic ability or aptitude, teacher rating scales, and measures of academic achievement (Callahan et al., 2017). Similarly, schools often make early algebra placements based on multiple years of prior test scores (e.g., Dougherty et al., 2017).

MAP Growth is an ideal measure of academic achievement to use as one of multiple data points at phase two given it is a computer adaptive test designed to measure academic achievement with precision even when a student is above grade-level. In contrast, many state summative achievement tests are not well-suited to measure above grade-level skills because they include few items assessing skills beyond grade level. Incorporating MAP Growth data can be useful if a district decides that a measure of academic achievement can help identify which students are ready for and would benefit from a service focused on developing skills that are more often taught at a higher grade.

A word of caution about "mastery"

MAP Growth is a measure of academic achievement and growth. As a computer adaptive test, it can provide precise measures of student achievement, even for students who are achieving above grade-level standards. However, it's important to distinguish a measure of achievement from a measure of content mastery. Because a second-grader taking MAP Growth might be posed with questions designed to measure skills typical of fourth grade content standards, MAP Growth shouldn't be thought of as a measure of grade-level content mastery because it adapts to the level at which the student is performing. That same second-grader might be answering questions on reasoning with geometric shapes (closer to fourth- or fifth- grade content), but she will not be answering as many questions on basic patterns and relationships (closer to second-grade content). This means it's possible for that high scoring second-grader to still have gaps in her learning of second-grade content. This should not be used as a reason to deny a student admittance in an advanced learning opportunity, but does indicate the need for identifying and filling in any of those foundational learning gaps. The more advanced the learning opportunity, the more forethought is necessary to make sure any gaps in essential learning have been or will be addressed.

Selecting measures for phase two

Before we dive into how, when, and why MAP Growth can be used as one data point when making service eligibility decisions, it's important to acknowledge that what services to offer in the first place is up to individual schools and districts. In some states, specific "gifted" services are mandated in all schools. But even if they are, schools should make service offering decisions based on the needs of their students. The goal is to make sure that all students are appropriately challenged so that they can continue to grow and develop. If some middle school students are so advanced in science that they need to have access to high school biology, the school should develop a process for making that service available. Similarly, if some students are so advanced in math that it makes more sense for them to attend math class with the students in a higher grade, then that should be an option. In general, student needs should drive the services that are offered.

In deciding which data points should contribute to placement decisions, schools should consider the principle of alignment, which poses the following question: How well aligned are the content, skills, and dispositions measured by a given assessment with those that will be fostered in the resulting service? The best data points for informing placement in a program or service are those that measure the prerequisite skills necessary to benefit from that program or service.[1]

Misalignment between what is measured at phase two and program content is likely to result in more false negatives (i.e., missing many students who would benefit from a program) and more false positives (i.e., recommending placement for students who won't end up being successful). Importantly, it is the *combined* phase-two criteria that need to be aligned with the service. Take reading ability as an example. A service might not be designed to foster advanced reading skills, but a certain level of reading proficiency might be required to benefit from the advanced class. For this reason, even an advanced or accelerated physics or chemistry class might benefit from having a measure of reading in phase two in addition to data points assessing content mastery of prerequisite skills in science and math. But educators should carefully consider what level of performance is needed. Students might only need grade-level reading achievement to benefit from an advanced physics class so that they can read course materials. Just because it's an advanced class does not mean students need be advanced in all domains at the same level.

The best way to judge alignment between the placement criteria and a program, service, or course is to review the content that assessment is designed to measure and compare it to the skills necessary to do well in the resulting service. MAP Growth measures student achievement in the areas of reading, mathematics, science, and language usage and is aligned to state standards in the states in which its partner schools operate.[2] MAP Growth scores are best positioned to support inferences about student learning and readiness for an advanced learning opportunity in those domains or that require those skills. There is wide variability in the focus of gifted and talented programs or accelerated math programs. Some offer enrichment of grade-level academic content while others focus on creativity, critical thinking, or leadership development. MAP Growth is less-well suited to inform placement decisions in nonacademic programs or services or those that don't require students to have

Just because it's an advanced class does not mean students need be advanced in all domains at the same level.

1 Note this is not the same as saying only students who have "mastered" prerequisite skills should be placed in a program or service. Students can have some degree of "skill gap" when it comes to grade-level material and still benefit from the advanced learning opportunity.

2 NWEA also publishes course-specific assessments for Algebra 1, Algebra 2, Geometry, integrated math courses, and biology. But as these are most-often given to students who are already enrolled in relevant courses, they are less useful for making placement decisions.

scored at high levels in the areas of math, reading, science, or language usage. Put differently, if a student's ability to benefit from an advanced learning opportunity does not depend on a certain level of math achievement (for example), then it doesn't make sense to condition participation on MAP Growth math scores.

Districts can also consider alignment empirically. If a school already has an identification system and service in place, it can and should review how well performance on the identification criteria correlates with success in the service. If students who do well on the actual placement criteria aren't doing very well in the course or program, it might be time to rethink the placement criteria. Similarly, if all students placed in a program appear to do equally well, it might be time to rethink the criteria. Perhaps an even wider group of students would benefit from and succeed in the service. For instance, imagine a scenario where students are only placed in an advanced and accelerated math class if they score at the 90th percentile on MAP Growth for math. After monitoring performance in the class, it becomes clear that many of those students are not successful. Maybe they are not performing well on quizzes or are frustrated with the depth or pace of the class. This pattern would indicate that this placement criteria on its own isn't well aligned with the service and there must be other skills or dispositions necessary to success that aren't being captured by high MAP Growth scores.

Combining multiple measures

When multiple data points are used to make eligibility decisions in phase two, the next question is how to combine them. After all, multiple data points will provide multiple pieces of information, but in the end, a dichotomous decision must be made as to whether a student should be placed in a service. There are three appropriately named ways to combine multiple data points: AND, OR, and MEAN combination rules.[3]

The **AND** combination rule can be thought of as the "multiple hurdles" rule because it requires a student to meet criteria 1 AND criteria 2 AND criteria 3 (and so on for as many criteria as a district requires) to be identified for a service. If a student misses just one of the criteria, they are not eligible. This rule makes the most sense when successfully meeting each of the criteria is necessary for success in the service. For example, if a student needs to have scored above-level in science and *math* to benefit from an above-level science class, an AND rule might be appropriate. Similarly, taking Advanced Placement U.S. History might require high scores in reading *and*

Put differently, if a student's ability to benefit from an advanced learning opportunity does not depend on a certain level of math achievement (for example), then it doesn't make sense to condition participation on MAP Growth math scores.

3 See Lakin (2018) and McBee et al. (2014) for more details on these rules and their implications.

language usage/writing. In these cases, both skills/scores are necessary to set a student up for success, so the AND rule makes sense. Other characteristics of AND rules are that, because of how restrictive they are, they will yield the smallest population of students with the most-homogenous level of skills, have the greatest opportunity for false negatives/missed students, and are the worst for equity. These drawbacks mostly stem from the "multiple hurdle" nature of AND rules. They're set up such that if you miss one hurdle, you're out. This is good in cases where jumping over every hurdle is necessary, but it's also a drawback because it means one mistake or bad testing day and you're similarly out. This is why schools should think carefully about whether the benefits of AND rules outweigh their characteristic drawbacks.

The **OR** combination rule is the opposite of the AND rule and can be thought of as the "multiple pathway" rule. Each additional assessment provides for a separate pathway to eligibility since a student only needs one qualifying score to be identified or placed in the service. Instead of needing high math AND reading scores, a student only needs high math OR reading scores. This rule makes the most sense for measures of similar content from different tests or different test events (e.g., fall OR winter). It makes less sense for data points that measure completely different skills (e.g., math OR reading) since it's hard to imagine a service for which high math scores OR high reading scores are indicative of readiness or need. Other characteristics of OR rules are that they will yield the largest population of students with the most variable level of skills, miss fewer students (fewer false negatives) but at the expense of accidentally identifying some (more false positives), and are the best for equity since fewer students in general are missed. By nature of providing many pathways to identification, this combination rule allows a student several opportunities to demonstrate a skill, often in different ways or at different times, and still be identified. Conceptually, OR rules just represent a lower barrier compared to AND rules.

Finally, the **MEAN** rule is the most complicated and can be thought of as the compromise rule because it falls between the AND or OR rule regarding size of population identified, variability of skill in those identified, false positives, false negatives, and equity. For example, MEAN rules will identify more students than AND rules, but not as many as OR rules (assuming the same number of data points). As with OR rules, MEAN rules make sense for multiple measures of the same or similar skills. For example, rather than requiring high reading scores AND verbal ability scores, taking the average makes more sense, particularly because the mean of multiple data points results in a higher overall system reliability—something desirable in any placement system. Important to the MEAN rule is that only assessments of similar content should be averaged together. Otherwise, the benefits described don't really work. For example, it makes sense to take the mean of MAP math scores from two testing occasions, but it doesn't make sense to average verbal ability and a math score. Of course, any time multiple scores on different scales are averaged, they must first be put on the same scale. For example, standardized verbal ability scores might have a mean of 100 compared to MAP, which is vertically scaled from 100 to 350. These two data points would need to be put in the same scale before they can be averaged.

When deciding how to combine multiple data points, a key factor should be the consequences of any incorrect decision. Put differently, what is the worst that could happen should a school place a student in a service even though they are not yet ready? If the risk of harm is high were a student to be placed in a service or course incorrectly, AND rules make more sense because the AND rule reduces false positives. There are relatively few instances of school-based programs where the risk of harm associated with false positives is high, but multi-grade acceleration (e.g., skipping a student from the end of tenth grade to college) might be a place where the AND rule is appropriate. If there is little risk of harm (e.g., students can be moved out of a program if it is not going well with little consequence), OR rules make more sense.

Schools often think about the risks associated with false positives but less often consider the risks of false negatives. Not allowing a student to move on when they are ready can stunt academic growth and result in boredom. Consider a student who is ready for algebra in seventh grade but is not allowed to take it until ninth grade. In a sense, this is a false negative. A service exists that would meet this student's needs, but the student wasn't allowed in. In this case the risks take the form of slower growth, less math learning over the scope of the student's K–12 career, and boredom. Schools need to weigh these risks appropriately. They do not want to place students in algebra who are not ready, but they also don't want to hold students back if they are. These considerations of the potential negative outcomes should be included when making decisions about cut scores or eligibility criteria more generally.

Summary of combination rules

AND

Definition: Must meet cutoff on all measures

- **Example:** All students at or above the 95th percentile in math and reading assessments are selected
- Useful in cases where meeting each of the criteria is necessary for success in the service

OR

Definition: Must meet cutoff on at least one measure

- **Example:** All students at or above the 95th percentile for fall or spring math scores are selected
- Useful in cases where the measures assess similar content

MEAN

Definition: Must meet cutoff on the average of all measures

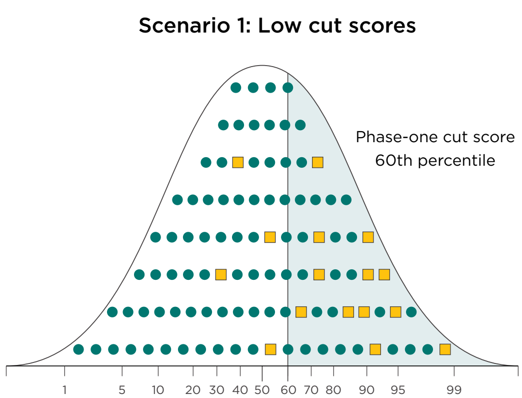
- **Example:** All students with combined average reading and verbal ability scores at or above the 95th percentile are selected
- Use to strike a balance between AND and OR rules

Factors to consider when setting cut scores

In any identification or program placement process, some criteria must be set above which students are placed in a service and below which they are not. Similarly, at the screening phase, some criteria must be established above which students move on to phase two and below which they do not. Where to place cut scores on the continuum of whatever is being measured at either phase is complicated. It is not possible to provide fixed recommendations given the multiple considerations at play that will vary from district to district. Instead, we offer guiding questions for districts to consider as they determine cut scores to use at phase one and phase two.

Phase-one cut scores

Setting cut scores involves balancing tradeoffs and depends on a district's priorities and financial constraints. The figure below depicts two different phase one scenarios to illustrate the tradeoffs of higher vs. lower cut scores. The curves depict the distribution of student test scores on a universal screener and their respective phase-one cut scores. Conceptually, some students score very low, some very high, and the majority are in the middle, but on any screener, most of the students who will end up being identified will be in the upper end of the distribution. Within the distribution, the yellow highlighted students depict "true positives"—students that are ready for and should be placed into the advanced learning opportunity. These are students the system should catch and send on to phase two. Of course, in the real world, we never know which students are true positives and even the best universal screener will miss some students and prevent them from being considered at phase two. This is because a single test score will never be a perfect approximation of a student's true ability. There are likely to be some students that would benefit from an advanced learning opportunity but are more challenging to identify because they scored below average on a screener for any number of reasons (e.g., they had a bad test day, they were assessed in their nonnative language which limits the test's ability to capture what the student knows, they are underchallenged in the classroom which leads to boredom and a tendency to disengage with instruction).



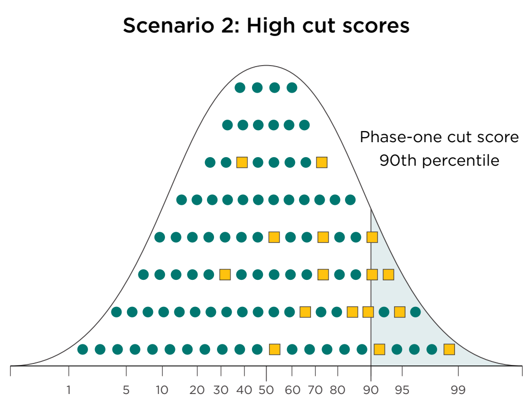


Figure 2. Tradeoffs of lower and higher cut scores in phase one

In the panel on the left, fewer students are missed (yellow students left of the shaded area) in phase two because a lower phase-one cut score means more students passed the screening phase. However, the trade-off is that many students are passed through to phase two who will not meet the eligibility criteria (the green students in the shaded area). This adds costs in terms of assessment dollars and student and staff time. In contrast, in the panel on the right, fewer resources are used because a higher phase-one cut score means fewer students are advanced to phase two, but as a result, many more students are missed (the yellow students who did not score in the 90th percentile on the screener).

Herein lies the challenge: schools want high sensitivity—they don't want to miss students who would benefit from a service—but they have finite resources and want to minimize unnecessary testing. When weighing the tradeoffs between missing students and conserving resources, schools should consider the associated equity implications. Research has shown that lower phase-one cut scores help make the identification process more equitable because more students from underrepresented backgrounds will move on to phase two and be identified (Card & Giuliano, 2016). Furthermore, lower cut scores also help catch students with disabilities or those who are learning English and, as a result, may score lower. The added costs associated with lower cut scores may be warranted to address persistent inequities within gifted programs and services.

Setting the phase-one cut score when a district wants to *equally* balance cost and sensitivity requires three pieces of information:

1. The reliability of the data at phase two
2. The correlation between the phases (also known as nomination validity)
3. The phase-two service eligibility cut score[4]

The added costs associated with lower cut scores may be warranted to address persistent inequities within gifted programs and services.

Under some reasonable assumptions (e.g., .95 reliability, nomination validity of .80), we can estimate that approximately three times as many students should pass through phase one as are to be identified at phase two. To concretize this, imagine a school sets a phase-two cut score at the 90th percentile such that 10% of students will be found eligible for the service (on average). A 90th percentile phase-two cut score necessitates a phase-one cut score at the 70th percentile (i.e., 30% of students move on to phase two). This assumes that the two phases are strongly correlated (i.e., at about .80) and that the reliability of those measures is high (i.e., at about .95). If either of those conditions are not met, the phase-one cut score must be lowered to avoid missing an unacceptably high number of students. Regarding MAP Growth,

4 See McBee et al. (2016) for technical details.

marginal reliability is often around .95 (NWEA, 2019). However, if other less-reliable data points are included in phase-two, it could bring the overall reliability down. Similarly, if MAP Growth is used as the phase-one universal screener and as one of the components in phase two, the nomination validity will be quite high. But both of these factors depend on the actual assessments being used at both phases. If that all seemed a bit too technical, McBee et al. (2016) give a detailed explanation of the statistical assumptions underlying this process and Peters et al. (2023) give a more conceptual explanation for how to set phase one cut scores.

Phase-two cut scores

When it comes to gifted and talented programs, many states prescribe the assessments and cut scores that lead to gifted service eligibility, most often based on national normative percentiles. For example, in Ohio, students are identified as gifted in a "specific academic area" if they score at or above the 95th percentile on an approved test in that domain (although the criteria for an accelerated or honors class can be different). Similarly, in Nevada, students must score at the 98th percentile on an individually administered test of cognitive ability. These two cases are somewhat atypical in that only one test score is required. Other states require multiple test scores. For instance, in Georgia, one pathway for gifted identification is a 96th percentile score on a "mental ability" test AND a 90th percentile score in a specific academic area (e.g., reading or math). And finally, there are also states like Illinois and New Jersey that specify local (i.e. school or district) norms should be used when setting phase-two cut scores. In these states, and others like them, the phase-two cut score is straightforward since all districts must use the criteria in state law.

If cut scores are *not* dictated at the state level, then districts should set them based on what level of prerequisite skill is necessary to benefit from the program (similar to college readiness cut scores) or they should set them based on the design of the program to challenge the most advanced learners in a given context under the premise that those students are the most likely to go underchallenged. For instance, is there a specific score on one or more phase-two assessments that is predictive of student success in the program? If so, that's where the eligibility cut score should be set. Or is the goal of the program to challenge the most advanced 5% of math students in the district? If so, then the district should set the cut score at the 95th percentile compared to all other students in the district (called a local norm). There is no one right or best criteria or cut score threshold that is applied to all programs, services, or across all states. That is why we recommend first consulting with any state rules or regulations and then considering the nature and purpose of the program or service when setting phase two cut scores or even selecting the data points used at phase two.

Conclusion

In the end, the goal of advanced program placement criteria is to correctly flag all students who are ready and would benefit so that they can receive a more appropriately challenging educational experience. Well-designed placement criteria can help schools challenge advanced learners in an equitable way by not missing students due to irrelevant factors such as lack of an initial referral. Universal screeners can correctly place more students in the resulting service with benefits falling disproportionately on students from traditionally underrepresented groups. The result is more students and more students of color and who are from low-income families enrolling in advanced learning opportunities (Card & Giuliano, 2016; Hyman, 2015; McBee et al., 2016). MAP Growth has many characteristics that make it attractive as a phase-one universal screener and as one component among several in phase two. The guidance laid out in this document can help schools design effective and defensible selection procedures for advanced learning opportunities.

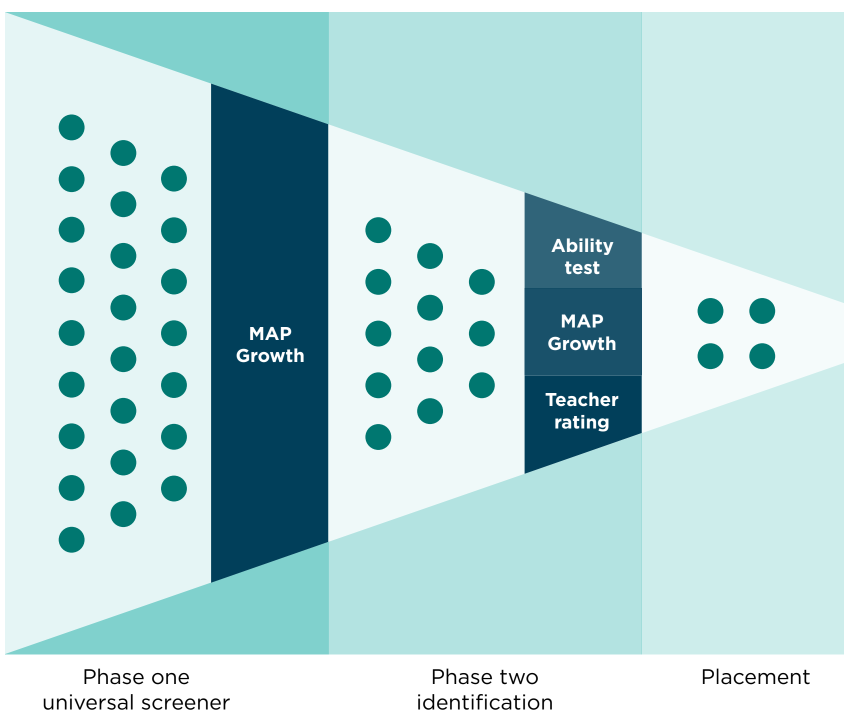
Examples

The following section gives three examples of how, when, and where MAP Growth could be used in making advanced learning placement decisions. This is certainly not an exhaustive list, but it applies to many of the concepts and criteria outlined above.

1. A two-phase gifted and talented identification system that balances cost and sensitivity

Several districts around the country make gifted and talented placement decisions based on three data points: a measure of academic achievement (like MAP Growth), a measure of academic ability or aptitude (like the Cognitive Abilities Test: Lohman, 2012), and a teacher rating scale of gifted characteristics or behaviors. For this example, we'll assume the district is only using the math portion of MAP Growth and the quantitative portion of the ability test along with the teacher rating scale to select second-grade students for placement in a third-grade compacted and accelerated math program, thus satisfying the basic requirements of alignment. We'll also assume the district uses building norms such that students are identified if they score in the top 10% (90th percentile within the school) of all second-grade students in their individual school.

As described above, a district could collect all three data points from all students in second grade and then combine them using the MEAN combination rule to make placement decisions. However, such a single-phase identification system would be costly. Instead, the district decides to use the math portion of MAP Growth, which is already administered to all students, as its phase-one universal screener. The figure below is a visual representation of this process.



This district sets its phase-two eligibility cut score at the 90th percentile within each school. This means that the top 10% of third-grade students, based on the average of the three data points from their second-grade scores, are placed in the compacted math program. However, the ability test data and teacher rating data would only be collected on students who scored at a certain level on MAP Growth for math. But how high must students score before those other two data points are collected (i.e., what is the phase-one cut score)? Earlier, we suggested roughly three times as many students should pass through the phase-one universal screener as the district wants to eventually identify. In this case, the district wants to identify 10% of students, which would set the phase-one cut score at the 70th percentile. Following these guidelines, the district could set its phase-one cut score at the 70th percentile and still identify 95% of the students they would have identified if they tested all students (universal consideration with all three data points).

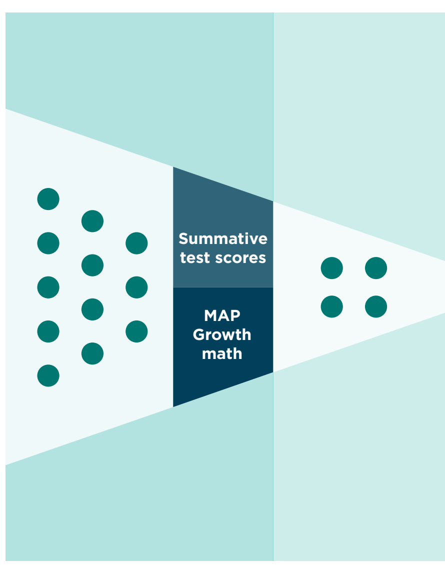
One challenge to this approach is that there is no way to identify which students would score in the top 10% on the average of all three data points when only some students are tested. In this school of 100 second-grade students, 30 will take the CogAT and be rated by their teachers because they scored at or above the 70th percentile in their school. So how is a school to know which of the 100 scored in the top 10 when only 30 students took two of the tests? The answer is strong nomination validity. Because MAP Growth for math (phase one) is going to be strongly correlated with the average of the three data points (phase two), we can be confident that the top 10 of 100 students are among the 30 who take the CogAT and are rated by their teachers. All a school would need to do is identify the highest-scoring 10 out of the 30 that took all three tests, and it could be confident that it’s identifying the top 10 out of their total class of 100. This is a little confusing because we’re no longer talking about percentiles—because we can’t know in what percentile a student scores unless all of the students in the norm group take all of the tests. But because we can safely assume strong nomination validity, we can be confident that the highest 10 are among the 30 students that passed through the universal screener and moved on to phase two.

Again, collecting ability, achievement, and teacher ratings on all students (a single-phase, universal consideration system) would result in higher sensitivity, but at far greater cost. Similarly, raising the phase-one cut score would decrease cost and involve less testing, but at the expense of missing more students. The 70th percentile MAP Growth score for math at phase one paired with the 90th percentile at phase two is a balanced approach in this example, but there’s no reason a district couldn’t err more on the side of cost savings (raising the phase-one cut score) or missing fewer students (lowering it).

2. Placement in seventh-grade algebra

Many districts are concerned about the equity of the advanced learning opportunities they offer. For example, nationally, gifted and talented student populations do not mirror the racial/ethnic demographics of the larger student population (Peters et al., 2019). And while they have similar rates of access at their schools, students of color have lower rates of enrollment in eighth-grade algebra than their white peers (Patrick et al., 2020). To try and address this problem and identify more students overall who might be ready for Algebra I in seventh grade, one district created a summative assessment to give to all its sixth-grade students at the end of the year. It covered skills the middle school math team felt were important to success in Algebra I (i.e., student understanding of ratios in solving real-world problems, unit rate problems involving constants, division of fractions, using variables to represent numbers and writing expressions, and problem-solving using four-quadrant planes). However, they were concerned that because it was so focused on specific pre-algebra skills, it might miss other important mathematical skills. The district also worried that while a student might be ready for algebra, he might have other gaps in his learning that they'd want him to cover first.

To address this concern, the district began including MAP Growth scores in math into their placement process. In addition to meeting the criteria on the sixth-grade summative assessment, students also needed to score at or above average for *eighth-graders* on MAP Growth according to NWEA norms (Thum & Kuhfeld, 2020)—an application of the AND combination rule. Allowing for score comparisons to above-grade students is a particular strength of MAP Growth. By including this data point, although students going into seventh-grade algebra might still have some skills they need to work on that would have been covered in standard grade seven math, they will have no greater degree of skill gap than their eighth-grade peers. The figure below presents this kind of multi-criteria, single-phase system visually. Because all data points used to make placement decisions are collected on all students already, there is no need for a two-phase process.



Single-phase identification system

However, were the district to decide that it no longer wanted to include the pre-algebra content on its sixth-grade summative assessment, it could move to a two-phase system. In that system, all students would complete MAP Growth, but only those students who scored at a certain level would move on to take the specific pre-algebra assessment. Because this is a hypothetical example, it's hard to know what phase-one cut scores would be optimal, but it's likely no more than 20% of students (those who score in the top 20% of MAP Growth math assessment according to national norms) would need to complete the pre-algebra assessment. This assumes that MAP Growth math scores and the district-created pre-algebra assessment are strongly correlated.

3. Full-grade acceleration

Grade-acceleration is the advanced learning intervention with the strongest research base and largest effects on subsequent student achievement (see Steenbergen-Hu et al., 2016). According to Hattie (2008), grade acceleration (e.g., skipping an advanced learner from the end of second grade to the beginning of fourth grade) has a larger effect on student learning than do many more-common educational interventions or practices such as direct instruction, enrichment programs, inquiry-based teaching, or collaborative learning. But the challenge is schools often struggle to decide which students should go through the in-depth eligibility determination process of deciding if grade acceleration is appropriate. It would be inefficient to conduct a child-study team meeting for every student in a grade on the question of whether full-grade acceleration is appropriate, just like not every student has an individualized education plan (IEP) as a result of a comprehensive evaluation with a school psychologist. What is needed is a universal screener to flag students who have a high likelihood of meeting the phase-two criteria for acceleration. This is an ideal scenario for using existing MAP Growth data—not to decide who should be accelerated, but rather to decide who should be considered for grade acceleration.

Most decision-making criteria for full-grade acceleration (see Integrated Acceleration System) are based on a points system and include factors such as academic achievement and ability, but also psychosocial and school factors. On the academic side, students receive the most "points" if they score at or above the 75th percentile for the grade level *two years above* their current grade (according to NWEA norms: Thum & Kuhfeld, 2020). For example, in fourth-grade math, the 75th percentile on MAP in the fall is a RIT score of 209—easily within the 99th percentile compared to second-grade norms. A second-grader scoring a 209 in math would receive the maximum points on the achievement and aptitude categories on the Integrated Acceleration System. A district could use MAP Growth with these cut scores at phase one to determine which students should go through the full Integrated Acceleration System. They are the most likely to do well since the school would already have evidence on hand that they would earn 100% of possible points in the "achievement" and "aptitude" sections of the Integrated Acceleration System. However, because points also come from other areas (such as school and family factors), these students still aren't guaranteed to be accelerated.

This process could be improved if the district in question also already collected academic ability data on all students in a given grade for other purposes (e.g., gifted and talented identification). If both those scores and scores from MAP Growth in math were available on all students, then those could be combined, via the AND rule, and used as a universal screener. Students who perform at the 75th percentile on the fall administration of MAP Growth math test for the grade level two years above their current grade AND who score in the top 10% (nationally) on an academic ability test would go on to be considered via the Integrated Acceleration System and a full child-study team. These students are even more likely to be strong candidates for grade acceleration than those who only have high MAP scores. Again, high performance on the two universal screeners would not 1) mean the students would always do well on the phase-two criteria or 2) end up being accelerated because there are other factors, such as whether the student wants to be accelerated, that go into such a decision. But alternatively, it's unlikely that any students who do not meet these screening criteria would have done well on the phase-two criteria. That's what makes for a good screener.

Frequently asked questions

1. **When using MAP Growth, which testing occasion should I use (fall, winter, or spring)? Similarly, should I look for a pattern of certain scores or is one score okay?** Although it might be attractive to look for a pattern of scores (e.g., a student must score at a high level in two of three testing occasions), doing so will exacerbate false negatives simply due to measurement error (i.e., a "pattern of scores" is the same as the AND rule where students must score at a certain level several times). Some districts will use a student's highest of three scores (an OR rule) while others will just use the score closest to when the program will start (e.g., spring scores for a program to start in the fall). Which to use probably depends on timing (i.e., districts may not have spring data in time to make placement decisions for next fall). There's no reason why one testing occasion is better than another aside from the logistics of when certain scores are available. The only suggestion we'd offer is to look at the past few testing occasions to make sure there are not obvious outliers. If the student's most recent score is also far lower than all others, a district might want to disregard that one on the basis that maybe it was a fluke occasion.

2. **Most of these recommendations seem to deal with talent in traditional academic areas of math and reading. What if my school wants to identify students in general intellectual ability or creativity?** In cases where MAP Growth doesn't measure the kinds of skills necessary for success in the program (such as with a creativity program), a district may still consider using it as a phase-one universal screener. The essential question is how strongly correlated MAP Growth scores are with the actual phase-two identification criteria. If they are strongly correlated, then using MAP Growth as a screener might make sense. If they are weakly correlated, MAP Growth could still be used, but the benefits would be lessened since the phase-one cut score would have to be set lower to compensate.

3. **Doesn't MAP Growth have ceiling effects that limit its use for gifted and talented identification?** No test can measure the entire range of possible math achievement. For example, MAP Growth is still designed around the Common Core State Standards or the standards of a given state. As a result, it doesn't measure math skills like calculus or skills that are not included in state content standards. However, because of its adaptive nature, MAP Growth can still yield precise scores even at very high levels of achievement. So while there is a ceiling, it is a very high ceiling.

4. **Are there set cut scores for programs like early algebra (subject acceleration) or early entrance to kindergarten?** Although there are linking studies showing what scores on MAP Growth are associated with college readiness or what scores on MAP Growth for math are associated with Algebra 1 proficiency, we do not recommend using a single cut score on a single test to make program placement decisions. Any single test score is always estimated with some degree of error, which is why strict cut scores (e.g., 200 you're in, 199 you're out) should be avoided.

5. **A lot of time is spent on the concept of alignment at phase two. What about alignment at phase one?** Alignment is most critical for phase two criteria because it is these criteria that decide who is placed in a particular service. Phase one criteria need only be correlated with the phase two criteria. The data points at the two phases do not need to be designed around the same underlying constructs as long as performance on phase one is strongly predictive of performance on the phase two criteria. This is because phase one isn't doing the identification. It's just helping decide who is going through the identification phase. Similarly, phase one criteria need not be aligned to any definition of "giftedness"—they need only be strongly correlated with the phase two criteria.

6. **Do the percentiles referenced here always refer to national norms?** What normative criteria should be used depends on the concept of alignment. As a general rule, the level of the normative criteria should match the level of the service. For instance, if a third-grade gifted and talented enrichment program is designed to challenge the most advanced students in a given school, then school norms make the most sense. Similarly, if the "program" is Advanced Placement Calculus, based on a national curriculum standard, then national normative criteria or achievement of prerequisite skills should be used as the criteria. We referred to national norms (often meaning MAP Growth norms) and local norms throughout the document as examples, and schools should follow the concept of alignment when deciding the actual program eligibility criteria (normative or otherwise). Everything described in this document could be applied to national or local norms. The only challenge comes when student scores are not normally distributed, which is more likely to be a problem when using smaller school-based norms. All of that said, which norms to use can also depend on state rules or laws since states like Illinois and New Jersey mandate school and district norms for gifted identification while Florida and Pennsylvania mandate national norms.

7. **Don't norms sometimes change or get updated?** Norms for all tests are regularly being refined and updated and those for MAP Growth are no different. When norms are updated, and when applied to historical data, that can mean that some students who were eligible before are no longer (or vice versa). In those cases, our recommendation is not necessarily to revisit decisions, but to always base decisions on the most recent norms available. When norms change, update policies, practices, or eligibility criteria for the next cycle. And in places where there is flexibility, always give benefit of the doubt to students.

8. **What effect did the COVID-19 pandemic have on gifted and talented student identification?** For most grades, test scores were lower in the years following the pandemic-related school closures than they were in years prior. Kuhfeld et al. (2022) reported that compared to fall 2019, fall 2021 test scores were .20 to .27 standard deviations lower in math and .09 to .18 standard deviations lower in reading. Although our work at NWEA (see Lewis et al., 2022) showed that students scoring in the top 10% were largely insulated from pandemic-related learning loss (i.e., mean scores for

the top 10% were similar), there is one important caveat. Because the rates of learning loss were different across groups, with students of color and those from low-income families being harmed more, these students are even less likely to score at high levels than they were before the pandemic (and they were already underrepresented). Any district that conducts identification for advanced learning opportunities post pandemic, but uses national normative percentiles based on pre-pandemic data, will see even greater racial, ethnic, and income disparities than they did before. Although this doesn't change the guidance in this document, it is something schools should be aware of and take steps to mitigate as part of their larger pandemic recovery efforts.

Glossary

Correlation: The relationship between two variables described on a scale of -1 (strong negative relationship) to +1 (strong positive relationship) with zero indicating no relationship.

Nomination validity: In concept, nomination validity is the degree to which the phase-one criteria are an appropriate metric for who should move on to phase two. As applied, nomination validity is the correlation between the two phases.

RIT score: The overall score for a subject based on a Rasch unit (RIT) scale that indicates how a child performed in a subject area.

False positive: A student was admitted to a program or service but shouldn't have been.

False negative: A student was denied admittance to a program or service but shouldn't have been.

Acceleration: An educational placement or intervention that involves a child moving through traditional curriculum earlier or at a faster rate than is typical.

Sensitivity: The proportion of students who would benefit from the program or service that are correctly placed in the service.

Universal screener: Any data point collected from all students in a grade for the purpose of determining which students should undergo further diagnostic assessments to determine program eligibility.

References

ACT. (2013). What are the ACT college readiness benchmarks? Information brief. https://www.act.org/content/dam/act/unsecured/documents/benchmarks.pdf

Callahan, C. M., Moon, T. R., & Oh, S. (2017). Describing the status of programs for the gifted: A call for action. *Journal for the Education of the Gifted*, *40*(1), 20–49. https://doi.org/10.1177/0162353216686215

Card, D., & Giuliano, L. (2016). Universal screening increases the representation of low-income and minority students in gifted education. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(48), 13678-13683. https://doi.org/10.1073/pnas.1605043113

Dougherty, S. M., Goodman, J. S., Hill, D. V., Litke, E. G., & Page, L. C. (2017). Objective course placement and college readiness: Evidence from targeted middle school math acceleration. *Economics of Education Review*, *58*, 141-161. https://doi.org/10.1016/j.econedurev.2017.04.002

Hattie, J. (2008). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Routledge.

Hyman, J. (2017). ACT for all: The effect of mandatory college entrance exams on postsecondary attainment and choice. *Education Finance and Policy*, *12*(3), 281-311. https://doi.org/10.1162/EDFP_a_00206

Kuhfeld, M., Soland, J., Lewis, K., Ruzek, E., & Johnson, A. (2022). The COVID-19 school year: Learning and recovery across 2020-2021. *AERA Open, 8*. https://doi.org/10.1177/23328584221099306

Lakin, J. M. (2018). Making the cut in gifted selection: Score combination rules and their impact on program diversity. *Gifted Child Quarterly, 62*(2), 210–219. https://doi.org/10.1177/0016986217752099

Lewis, K., Kuhfeld, M., Langi, M., Peters, S. J., & Fahle, E. (2022). The widening achievement divide during COVID-19. https://www.nwea.org/uploads/2022/11/CSSP-Brief_Widening-achievement-divide-COVID-19.pdf

Lohman, D. F. (2012). *Cognitive Abilities Test Form 7 research and development guide*. Riverside Publishing.

McBee, M. T., Peters, S. J., & Miller, E. M. (2016). The impact of the nomination stage on gifted program identification: A comprehensive psychometric analysis. *Gifted Child Quarterly*, *60*(4), 258–278. https://doi.org/10.1177/0016986216656256

McBee, M. T., Peters, S. J., & Waterman, C. (2014). Combining scores in multiple-criteria assessment systems: The impact of combination rule. *Gifted Child Quarterly*, *58*(1), 69–89. https://doi.org/10.1177/0016986213513794

NWEA. (2019). MAP Growth technical report. Portland, OR: Author. https://www.nwea.org/uploads/2021/11/MAP Growth-Technical-Report-2019_NWEA.pdf

Patrick, K., Socol, A., & Morgan, I. (2020). Inequities in advanced coursework. https://edtrust.org/resource/inequities-in-advanced-coursework/

Peters, S. J., Gentry, M., Whiting, G. W., & McBee, M. T. (2019). Who gets served in gifted education? Demographic representation and a call for action. *Gifted Child Quarterly*, *63*(4), 273–287. https://doi.org/10.1177/0016986219833738

Peters, S. J., Makel, M. C., Lee, L. E., Stambaugh, T., McBee, M., McCoach, D. B., & Johnson, K. (2023). What makes for an effective gifted and talented screener? https://doi.org/10.35542/osf.io/sqkae

Peters, S. J., Stambaugh, T., Makel, M. C., Lee, L. E., McBee, M. T., McCoach, D. B., & Johnson, K. R. (2023). The CASA criteria for evaluating gifted and talented identification systems: Cost, alignment, sensitivity, and access. *Gifted Child Quarterly*, *67*(2), 137–150. https://doi.org/10.1177/00169862221124887

Steenbergen-Hu, S., Makel, M. C., & Olszewski-Kubilius, P. (2016). What one hundred years of research says about the effects of ability grouping and acceleration on K–12 students' academic achievement: Findings of two second-order meta-analyses. *Review of Educational Research*, *86*(4), 849–899. https://doi.org/10.3102/0034654316675417

Thum, Y. M., & Kuhfeld, M. (2020). NWEA 2020 MAP Growth achievement status and growth norms tables for students and schools. NWEA Research Report. Portland, OR: NWEA. https://teach.mapnwea.org/impl/NormsTables.pdf

About the author

Scott J. Peters specializes in educational assessment and data use, gifted and talented student identification, equity within advanced educational opportunities, and effectiveness of educational policy. His research focuses on how schools can leverage assessment data for maximum school and student benefit. His ongoing projects relate to balancing cost, sensitivity, and equity in gifted and talented student identification, how to proactively screen students for advanced learning opportunities, examining growth trajectories for advanced learners, and how to ensure all students have access to advanced learning opportunities.

Prior to coming to NWEA, for 13 years, Dr. Peters served as a professor of assessment and research methodology at the University of Wisconsin—Whitewater. His scholarly work has appeared in the *Australian Educational Researcher*, *AERA Open*, *Teaching for High Potential*, the *British Journal of Educational Psychology*, Exceptional Children, *Gifted Child Quarterly*, and many other publications. He received his PhD from Purdue University in educational psychology and applied research methodology.

Acknowledgments

This paper underwent several rounds of internal and external review to help strengthen our recommendations and guidance. The following people provided generous feedback to this document, and we thank them for their time and effort in helping us improve and refine our guidance.

Melissa Hinshaw, Coordinator of Assessment, Center for Talent Development, Northwestern University

Darin Kelberlau, Executive Director of Assessment, Millard Public Schools, NE

Benjamin Lacina, Talent Development and Advanced Academics Coordinator, South Washington County Schools, MN

Kim Rensch, Gifted and Talented Specialist, Fargo Public Schools, SD

Meredith Langi, Research Associate/Psychometrician, WestEd

Karyn Lewis, Director, Center for School and Student Progress, NWEA

Monica Simonds, Director, Advanced Learning Programs and Services, Richardson ISD, TX

Susie Spafford, Sr. Account Manager, Partner Accounts, NWEA

Sylvia St. Cyr, Senior Director, Partner Accounts, NWEA

Matthew Makel, Professor and Chair of High Ability Studies, University of Calgary

Colleen Boyle, Curriculum Coordinator, Gahanna-Jefferson Schools, OH

Matthew T. McBee, Vice President of Data Science, Service Management Group

SEP23 | WELTSK6662