

NWEA response to Zavitkovsky paper

The following brief was developed by NWEA® in response to Paul Zavitkovsky’s paper titled [“Data Rich/Information Poor.”](#) which was presented at NCME on April 14th, 2023. This brief includes high-level comments and concerns regarding Mr. Zavitkovsky’s paper—which has not yet been subjected to a formal peer review publication process—and is not meant to be an exhaustive, line-by-line critique.

What follows are five areas that challenge the strength, validity, and quality of some of Mr. Zavitkovsky’s findings, and as such, provide context for interpreting his results with caution and scrutiny..

Inappropriate comparisons

In his paper, Mr. Zavitkovsky makes several inappropriate comparisons of results across different tests and groups of students. For example, in his “Case #3,” the author cites a comparison of MAP® Growth™ “proficiency” estimates for students in the spring of eighth grade in District 65 to grade equivalent scores for ninth grade students in the fall on the STAR 360 Screening test in District 202. This means he is comparing the results of two different groups of students in two independent school districts at two different time points across two different assessments using two different outcome measures. It is unclear why we would expect these results to be even remotely comparable.

To provide another example, in Figure 2.3, the author compares the percentage of students at or above the 50th percentile based on results from NWEA’s nationally representative 2015 norms to PARCC’s 2015 Illinois norms. The author suggests that these approaches are comparable (referencing Illinois as “the most demographically representative state”) while failing to acknowledge that these comparisons are based on two different sets of norms which use two different norming approaches from two different tests at two different levels of aggregation (i.e. national vs. Illinois). It is again unclear why these comparisons would be expected to yield similar results.

Lack of comparability of samples

Throughout the paper, the author makes comparisons of test results across many different groups of students but fails to provide any data on the comparability (or lack thereof) of the different student groups. For example, in Figures 2.1 and 2.3, the author does not include any information on the number of students tested, what proportion of the district tested on the different assessments, the demographics of the different student groups, or any indication for how representative or comparable the different groups of students might be. These tables also do not include information on policies that may govern which students in a district were administered the different assessments, which is important context especially given that MAP Growth is an assessment of choice, whereas a test like PARCC is not. Absent any information on the comparability of samples—a standard practice for peer-reviewed publications—it is further unclear why we would expect these results to be in any way comparable. This lack of comparability significantly limits the interpretation of any of the presented differences across student groups.

Incorrect definitions and uses of data

“Proficiency” in this paper appears to be defined as students achieving at the 50th percentile relative to NWEA’s nationally representative MAP Growth norms. However, the MAP Growth assessment was not designed to measure student proficiency, nor is the 50th percentile a threshold NWEA would recommend be used as a proxy for “proficiency.” Instead, NWEA conducts linking studies to estimate MAP Growth RIT scores that predict proficient performance on state summative assessments. These linking studies generally indicate that students need to achieve much higher than the 50th percentile to be on track to be proficient on these end-of-year tests. Thus, the way “proficiency” is defined and used in this paper is not correct and is not an approach NWEA recommends.

The author also translates MAP Growth results into “grade equivalent” metrics at several points throughout his paper, including in Figures 2.1 and 2.2. However, he provides no supporting information about how he calculated these grade equivalent metrics, which limits a reader’s ability to validate the author’s approach to these calculations. Beyond that, NWEA neither provides nor recommends that MAP Growth results be translated to this type of grade equivalent metric and does not support these types of translations.

Misrepresentations of national trends

The author notes the following on page 6—“Given the sustained attention that NWEA had paid to developing highly national norms in 2011 and 2015, large differences in results reported using 2015 NWEA and NAEP norms were surprising.” The author then proceeds to support this claim using data from the Chicago Public Schools.

For information about how national MAP Growth and NAEP trends compare, the author could have simply referenced NWEA’s 2020 norming study, which includes this exact information. The figure below shows that national trends in fourth- and eighth-grade math and reading follow a similar pattern across both assessments.

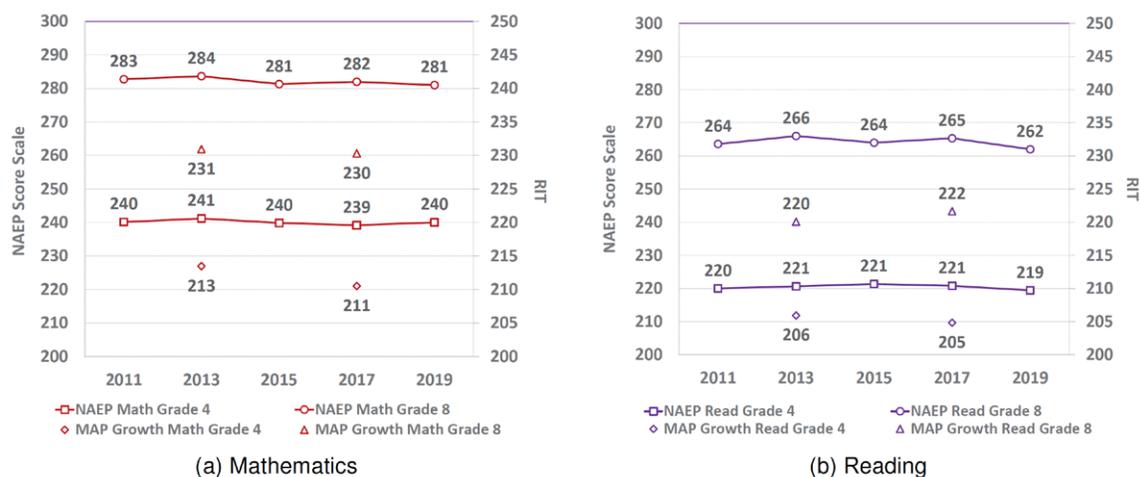


Figure 6.4: NAEP and MAP Growth trends for grades 4 and 8 mathematics and reading

Thum, Y.M., & Kuhfeld, M. (2020). *NWEA 2020 MAP Growth Achievement Status and Growth Norms for Students and Schools*. NWEA Research Report. Portland, OR: NWEA

Misrepresentations and inaccuracies

Throughout his paper, the author makes a number of misrepresentations, inaccuracies, and unsubstantiated claims. On page 3, the author notes:

“These anomalies were especially noteworthy because they arose during a period when NWEA technical manuals indicated that the organization was intensifying its efforts to report accurate, nationally representative norms from its large but not-fully-representative national testing sample.”

In support of this sentence, the author cites NWEA’s 2011 RIT Scale Norms Study, while failing to note that both the 2015 and 2020 norming studies discuss the methodology NWEA employs to sculpt its robust set of nationally representative norms, a subject that is also discussed in NWEA’s 2019 MAP Growth Technical Manual.

In Figure 1.2, and at other points throughout the paper, the author presents achievement trends on a number of different assessments, including MAP Growth, to provide some indication of “anomalous trends” in eighth grade math. However, in those graphs, he uses different y-axes across the assessments, making a 7-point change on MAP look greater than an 11-point change on NAEP-TUDA, all while putting ISAT scores (ranging from 259-268) and PARCC scores (ranging from 721 to 717) all on the same line/scale.

In that same figure, the author also shows a flat trendline on MAP Growth for all eighth-grade US public school students, with a presented score of 231 in both spring of 2013 and 2019. In fact, NWEA’s 2015 and 2020 norming studies show that the eighth-grade norm in mathematics in the spring that would most closely approximate the listed time points as scores of 231 and 230 respectively. This slight decline in mathematics achievement nationally is also reflected in both the PARCC and national NAEP trends the author presents at the bottom of this figure—a consistency in national and state trends the author fails to mention at any point in the paper. In fact, on page 3, the author notes that “problems arose from NWEA scoring trends because they were substantially more positive than comparable reportage from NAEP and other large-scale systems.” This statement is not supported by the data included in Figure 1.2, especially when the correct MAP Growth scores are included.

Finally, and most critically, the author notes in his conclusion of the need to “take NWEA to task for knowingly misrepresenting interim assessment as a necessary and widely-acknowledged ingredient of comprehensive assessment systems.” NWEA did not misrepresent anything and instead offered its perspective on how interim assessment could be used in conjunction with other assessments. While other groups may disagree with that statement—as the author of this paper clearly does—it is NWEA’s belief that interim assessments, when used correctly, can serve a critical role in the comprehensive assessment framework utilized in a school or district.



NWEA, a division of HMH, supports students and educators worldwide by providing assessment solutions, insightful reports, professional learning offerings, and research services. Visit [NWEA.org](https://www.nwea.org) to find out how NWEA can partner with you to help all kids learn.

© 2023 NWEA. NWEA and MAP are registered trademarks, and MAP Growth is a trademark of NWEA in the US and in other countries.

MAY23 | WELTSK6109