

thinkpoint

Online assessment in 2020-21: Putting the results in context and maximizing their utility

By James A. Gullen, with Dodie Raycraft, and Ben Allen

Introduction

In 2020 and 2021, U.S. educators were asked to adapt in unprecedented ways in response to the COVID-19 pandemic that swept the nation. Public health mitigation measures required schools to shut down or cycle through a variety of learning models:

- Widespread distance learning
- Hybrid models that included both distance and face-to-face learning
- In-person learning with changes to the school environment and procedures to mitigate the spread of infection

Given the unpredictability and severity of the pandemic's effects, it was often not possible to provide adequate planning time to thoroughly understand all of the implications of these quickly enacted changes, which impacted every aspect of K-12 education.

Online assessment is one area that clearly demonstrates a significant challenge posed by the pandemic for public education. In Michigan, schools are required to administer interim-benchmark assessments. Many schools have chosen to administer these assessments online in an effort to maximize the efficiency of giving these tests while also minimizing down-time from instruction.

Given the context of the pandemic, administration has varied building-to-building with some assessments administered in the school building and some administered remotely in homes or other sites. Mon-



itoring or proctoring the assessment administration also differed district to district and perhaps even school to school. Like summative assessments (i.e., assessments of learning), interim-benchmark assessments are typically required to be administered under standardized conditions, so the variability in assessment administration due to the required public health measures may have had an impact on utility of results from these assessments.

This standardization of testing conditions is what allows for the valid comparison of scores between classrooms, schools, and districts, as well as comparisons over time. To the extent that testing conditions are not standardized, (such as some stu-

dents receiving assistance in taking the tests), comparisons may lose their accuracy and meaning. With all of the changes to public education necessitated by the COVID-19 virus and the subsequent variability in administration procedures and conditions, legitimate questions are being raised about the utility of scores generated from interim-benchmark and summative assessments that are administered under differing conditions.

Given that research on the impact of public health COVID mitigation measures is just now starting to be addressed, some important research questions for consideration could include:

- What impact, if any, does online remote administration of

interim-benchmark and/or summative assessment have compared to online in-school, administration?

- Can results from the same online interim-benchmark assessment administered to students in school be compared to those of students who took the online test remotely?
- Can results from online assessments administered in previous years, in school, be legitimately compared to results acquired by remote administration of the same tests this year?
- Are there effective methods to mitigate inappropriate test-taking activity on the part of test-takers in an online administration environment? (See Li, M., et. al., for examples of research currently going on.)

Early research suggests the scores generated from remote administration are comparable to results obtained from in-person administration, both in the same school year and previous school years. This same research also shows, however, that there are differences in critical aspects of assessment between remote administration and in-person administration. Student engagement with the assessment, which we know affects student performance on a test, is one area that shows differences between administration modalities. (Kuhfeld, et.al., 2020) Clearly, there is a need for more research into these issues.

While research into these important issues will be conducted, it will not happen overnight. In the meantime, there are things that schools and districts can do, themselves, to understand and improve the quality of their test scores.

What to Do:

With all the remote and hybrid instructional models students have experienced, it would be prudent to evaluate the scores obtained from students during the pandemic. Essentially, COVID-19 countermeasures have created new demographic categories for which we can observe data.

The simplest study would be comparing scores between students who completed an interim-benchmark or summative assessment online remotely and those who completed it online in person. Do we see any overall differences between these two groups? If we do see a difference, more inquiry may be needed. Is there a reason, other than administration method, that might account for this difference?

Looking back to previous years' scores might also be warranted. Do we see this difference in scores



between these groups of students (those who tested online remotely and those who tested online in person) before administration conditions changed? For example, if we look at the aggregate scores on the assessments given in the 2019-2020 school year for these same students, is the difference, if there is any, consistent with what was seen during the pandemic? Pursuing those answers give insight into the nature of the data, how it may and may not be interpreted, and what those data should and shouldn't be used for.

It is easy to see that there are a number of relevant research questions that a school or district can ask in this vein. Addressing those questions in a thoughtful manner will help educators understand the data that we currently have and how trustworthy it is.

How can we maximize the quality of the data that we will get in the near future?

To maximize the quality of data obtained from tests designed to be administered in standardized conditions, work is necessary to eliminate—or at least minimize—differences in administration. For online assessments, this includes technical measures such as ensuring that students participating remotely have bandwidth comparable to that which is available to students taking

the test in school. Access to technical support for remote test takers must also be available and convenient. The key is to work to create an equitable testing situation for all students to the greatest extent possible.

Communication with parents of students who participate remotely will be critical to identify equity issues, explain testing requirements, and provide support. The school should communicate the desired administration conditions and why those conditions are important. Clear guidance in how parents and other family members can appropriately support their student in the testing process will help minimize the differences in administration conditions.

This would also be a fine time to help families understand the purpose of these assessments in helping to monitor students' progress and adjust the instructional program to support learning. For example, parents and siblings should refrain from helping the student answer test questions, since this may provide a misleading picture of the student's current level of achievement. Helping them to understand how interim-benchmark and summative assessments fit into a balanced assessment system in particular and a sound educational program in general would also be helpful.

Summary

In summary, by taking a retrospective look at the data obtained by mixed-format online assessment administration, schools and districts can gain a better understanding of the limitations of their data. By implementing a robust communication plan with parents of students who test remotely, schools and districts can improve the quality of their online assessment results going forward.

Although research into these important areas will be carried out, a school or district doesn't have to wait for those results and findings to move forward. They can, and should, work to move forward themselves through reflection and analysis to determine their "next steps."

References

- Kuhfeld, M., Lewis, K., Meyer, P. & Tarasawa, B. (2020). Comparability analysis of remote and in-person MAP growth testing in fall 2020. NWEA Technical Brief. <https://www.nwea.org/content/uploads/2020/11/Technical-brief-Comparability-analysis-of-remote-and-inperson-MAP-Growth-testing-in-fall-2020-NOV2020.pdf> Accessed April 17, 2021.
- Li, M., Luo, L., Sikdar, S. et al. Optimized collusion prevention for online exams during social distancing. *npj Sci. Learn.* 6, 5 (2021). <https://doi.org/10.1038/s41539-020-00083-3>. Accessed April 26, 2021.



JAMES G. GULLEN, Ph.D., has an extensive education background, serving as an educator in diverse settings, including local and intermediate school districts and the Michigan Department of Education. He is especially skilled in helping educators understand technical measurement concepts and develop a critical view of mandated uses of data. Jim has earned

a Bachelor of Science degree in mathematics education, and M.A. and Ph.D. degrees in educational evaluation and research, all from Wayne State University.

Dodie Raycraft, Educational Consultant for the Michigan Assessment Consortium, and Benjamin Allen, Detroit Public Schools Community District, contributed to this paper.

This ThinkPoint was developed through the Michigan Assessment Literacy Facilitators Association (MALFA), a MAC professional community that promotes collaboration among those charged with helping educators develop, select, and use high-quality assessment, as well as use the results from these assessments to improve instruction and learning.