VIA HAND DELIVERY, U.S. MAIL and E-MAIL

Tymica Dunn
Procurement Manager
North Carolina Department of Public Instruction
(301 N. Wilmington St, Room B04)
6301 Mail Service Center
Raleigh, NC 27699-6301
tymica.dunn@dpi.nc.gov

       Re: *RFP No. 40-RQ20680730: Read to Achieve Diagnostics – Software As A Service*

           Protest Letter and Request for Protest Meeting

Dear Ms. Dunn:

       Pursuant to Section D(8) of the Read to Achieve RFP referenced above and issued by the Department of Public Instruction ("DPI"), and pursuant to 09 NCAC 06B.1102, we hereby timely submit this protest letter and request for a protest meeting on behalf of Amplify Education, Inc. ("Amplify"). We also request that the State suspend or terminate performance of the contract awarded to Imagination Station, Inc. ("IStation") on June 7, 2019 during this protest.

       As explained below, IStation's product plainly fails to meet key requirements established by North Carolina law, including in demonstrating accuracy in assessment outcomes, and in identifying students with dyslexia. Further, IStation's ISIP product does not use "developmentally appropriate practices" for the K-3 population as required by statute, and will not deliver the teacher interaction needed to properly assess student reading skills or capture the detailed data teachers rely on to deliver effective reading instruction in elementary classrooms. In addition, IStation's assessment does not satisfy express requirements of the procurement at issue in this letter. Because it does not meet North Carolina's needs and requirements, the use of IStation will lead to outcomes that are detrimental to North Carolina students.

       As many who participated in this evaluation have now stated publicly, Amplify was the most qualified bidder and should have been awarded the contract. Its solution would be the best value to the State of North Carolina. Indeed, Amplify understands this was the recommendation of the evaluation committee for this procurement, which specifically identified many of the problems with IStation's offering which are discussed in this letter.

This protest letter is based on information currently made available to Amplify. Amplify has outstanding public record requests for the evaluation record, awarded contract, and related documents for this procurement. Amplify reserves the right to supplement this letter with additional information.

Amplify appreciates the opportunity to raise its legal and practical concerns to DPI, and values its existing long-term partnership with DPI.

I. Background

A. Amplify

Amplify was founded in 2000 and has 17 years of expertise in K-12 education and in technology providing data, technology-supported teacher tools, and instructional resources that help teachers generate student success. Amplify provides services to school districts in all 50 states, and has been pleased to partner with North Carolina for almost ten years to help educators teach children to read by third grade. Amplify has provided K-3 formative and diagnostic literacy assessment to North Carolina schools since 2005, and has been a provider to the State of North Carolina since 2010.

B. North Carolina Initiatives and Read to Achieve

North Carolina has a long history of commitment to the selection and use of developmentally appropriate K-3 reading assessments. More specific to North Carolina's current program, in 2012, the North Carolina General Assembly adopted legislation for the North Carolina Department of Public Instruction to implement the Read to Achieve ("RtA") program.

The right to a "sound basic education" is a constitutional right in North Carolina, *Leandro v. State*, 346 N.C. 336 (1997). Reading is the most fundamental aspect of basic education, and to that end, the General Assembly has adopted RtA legislation to ensure all North Carolina students can read. Under N.C. Gen. Stat. § 115C-83.1, it is the public policy of the State "to ensure that every student read at or above grade level by the end of third grade and continue to progress in reading proficiency so that he or she can read, comprehend, integrate, and apply complex texts needed for secondary education and career success." Thus, the legislation is intended "to ensure that . . . difficulty with reading development is identified as early as possible," and that "students receive appropriate instructional and support services to address difficulty with reading development and to remediate reading deficiencies." N.C. Gen. Stat. § 115C-83.2(a)(i).

A core component of the Read to Achieve program is "universal screening." Universal screening means the early assessment of all students, with follow-up assessments to monitor progress, as a way to determine which students are at risk for reading difficulties and to determine

how educators should respond to improve student outcomes. North Carolina's RtA law requires the State to provide schools with "valid, reliable, formative, and diagnostic reading assessments." N.C. Gen. Stat. § 115C-83.6(a). These assessments must use "developmentally appropriate practices" to "address oral language, phonological and phonemic awareness, phonics, vocabulary, fluency, and comprehension." *Id.* § 115C-83.6(b).

In addition to these general legislative provisions on universal screening, the General Assembly passed a statute in 2017 that focused specifically on screening for dyslexia and other learning difficulties. 2017 N.C. Sess. Laws 127 (**Exhibit A**). This law requires "that all students with specific learning disabilities, including dyslexia . . . receive the necessary and appropriate screenings, assessments, and special education services to provide interventions for learning difficulties." *Id.* § 1. The statute requires the provision of ongoing professional development for teachers on identifying dyslexia and intervening to improve outcomes. *Id.* § 2. The statute also requires local boards of education to "review the diagnostic tools and screening instruments used for dyslexia . . . to ensure that they are age-appropriate and effective," and to "determine if additional diagnostic and screening tools are needed." *Id.* § 4.

C. Procurement History

1. Prior Contract and Recent RFP

Amplify was awarded the most recent statewide "Read to Achieve Diagnostics – Software as a Service" contract after a competitive procurement in 2016. DPI has exercised both of its option years, and the contact is now set to expire on August 24, 2019.

In the fall of 2017, DPI issued a new RFP for these services. However, the RFP was cancelled on March 22, 2018. A primary stated reason for the cancellation was the need to "begin again with clarified specifications and additional time for implementation." (**Exhibit B,** March 22, 2018 email from Michael Beaver of DPI to Amplify). Importantly, the cancellation recognized that a fair competition would require sufficient time for implementation, regardless of what bidder prevailed in the procurement.

On September 6, 2018, DPI issued the RFP at issue here. (**Exhibit C**, RFP No. 40-RQ20680730, the "RFP")). Though the RFP did not include a specific deadline for a final award, it plainly anticipated an award around the end of 2018. Vendor submissions were due on October 1, 2018, and vendor demonstrations would be held on October 22-23. After the vendor demonstrations, DPI requested clarifications on two additional occasions, which Amplify provided by required deadlines.

2. Key RFP Requirements

Procurement Manager
North Carolina Department of Public Instruction
June 24, 2019
Page 4

The RFP contains mandatory requirements for bidders to include in their proposed solutions, all of which are dictated by the requirements of the RtA legislation and the General Assembly's 2017 dyslexia statute. Thus, the RFP requires the vendors' proposed solutions to include the following:

- Deliver "adequate classification accuracy," *id.* Attachment A, Spec. 4, at p. 25;

- "[A]dequately identify indicators of risk for dyslexia in grades K-3," *id.* Attachment A, Spec. 8, at p. 25; and

- "[A]ssess student progress," using monitoring tools that are "brief," "repeatable," and "sensitive to improvement over time, including short-term change," id., Section I ("Introduction"), at p. 6; *id.* Attachment A, Spec. 6, at p.25

- "[D]irectly" assess "reading and pre-reading behaviors", including "oral language (expressive and receptive)" and "accuracy and rate" *See* RFP, Attachment A, Spec. 1 and 2, **Exhibit C**, at 25.

The key policy underlying the RFP, the RtA statute, and any universal screening program aimed at identifying at-risk students is that a solution must provide "developmentally appropriate practices to assess K-3 students," RFP, Attachment A, Spec. 9, **Exhibit C** at p. 26; see N.C. Gen. Stat. § 115C.83.6(b). The definition of "developmentally appropriate" must necessarily vary based on the age of the students, and a clear standard has developed as to the attributes of what is "developmentally appropriate" for students in the early years of elementary school. Many of those attributes can exist only if the reading assessment involves a teacher observing, listening and guiding a student through tasks requiring a student to read aloud or produce sounds and words demonstrating proficiency in key measured skills. Examples of these attributes include:

- Assessments for young children should consider students' "performance on authentic activities" rather than artificial exercises.[1] Put simply, in order to understand if students know their letters, letter sounds, or can read, the mCLASS assessments have students perform these tasks directly. Reading out

---

[1]Nat'l Ass'n for the Educ. of Young Children, Position Statement: Developmentally Appropriate Practice in Early Childhood Programs Serving Children from Birth Through Age 8 (2009), at 9, https://www.naeyc.org/sites/default/files/globally-shared/downloads/PDFs/resources/position-statements/PSDAP.pdf (in determining what practices are developmentally appropriate, it is important to consider"[w]hat is known about child development and learning—referring to knowledge of age-related characteristics that permits general predictions about what experiences are likely to best promote children's learning and development").

loud is an example of an authentic activity akin to one children engage in every day in their classrooms and at home with their parents or guardians. On the other hand, stopping reading in order to select from a drop-down menu of multiple choice prompts (as IStation does) is not a type of reading that children or adults will encounter in any other context.

- Special care must be taken to ensure all students of this age understand the instructions for the assessment.[2] Even for young children familiar with computer devices and having the manual dexterity to use a mouse, following prompts from a computer while figuring out the user interface may be bewildering, compromising the results of the assessment. Contrast that with the experience of a familiar adult guiding the child through the assessment, recognizing and adjusting for signs of confusion in understanding the instructions.

- To ensure accuracy, the assessment must include procedures with standard accommodations for teachers to apply professional judgment in ensuring that students demonstrate their abilities to the fullest extent. For example, a teacher assesses oral fluency by determining how many words a student can accurately read within a defined period of time. Using Amplify's mCLASS, a teacher is trained to prompt a student to move past a word if the student is unable to recognize the word within three seconds, or, if a student seems distracted by some other event occurring in the classroom or out the window, a teacher can exercise judgment in bringing the student's attention back to the task.[3]

Recognizing that live teacher observation and engagement is a prerequisite component of a developmentally appropriate literacy assessment tool, the RFP asks vendors to describe several aspects of their products' "observation-based practices." RFP, Attachment A, Spec. 1, 2, **Exhibit C** at p. 25.

Four bidders submitted bids, including Amplify and IStation. The details of the Amplify and IStation proposals, including the failure of IStation's proposal to provide a developmentally appropriate literacy assessment tool to meet the requirements of the RFP and North Carolina law, will be addressed below.

---

[2] *See* Am. Educ. Research Ass'n, Standards for Educational and Psychological Testing (2014) ("AERA Standards").

[3] The AERA Standards for Educational and Psychological Testing explain that the usefulness and interpretability of test scores require that the directions to examinees be standardized and that these directions be understood fully by all examinees. *See* AERA Standards. The interactions between teacher and student during observational assessment can ensure that this takes place especially for young students in kindergarten through third grade.

3. Cancellation and Award

On February 21, 2019, DPI issued a letter announcing the cancellation of the procurement. (**Exhibit D**, February 21, 2019 Cancellation Letter.) Under DIT Rule 09 NCAC 06B.0401, an agency may cancel a procurement on specific grounds. However, the Cancellation Letter does not identify any grounds for the cancellation of the RFP here. To date, Amplify has not received any agency record, as required by the Public Records Law and the rules of DIT, explaining why the bid has been cancelled.

Pursuant to the Cancellation Letter, DPI then entered into negotiations with bidders under 09 NCAC 06B.0316. We understand that DPI conducted negotiations with two bidders, Amplify and IStation. On April 11, 2019, Amplify had an in-person negotiation session with DPI and following that meeting provided written clarifications and Alternate Cost Responses, as requested by DPI. After the final clarifications were submitted on April 23, 2019, Amplify received no further information from DPI.

On June 7, 2019, Amplify learned that Superintendent Johnson had issued a letter to school districts announcing that IStation had just been awarded the contract. (**Exhibit E**, June 7, 2019 Johnson letter). The award announcement on June 7—the last day of the school year in many districts—allows little time for schools to implement IStation's product and train teachers on how to effectively use it.

It has been reported that IStation's contract is for $2.8 million a year, or $8.3 million over three years. While DPI informed the press that Amplify's current contract is for $6.3 million a year, it apparently did not disclose that Amplify's bid on this contract was $3,755,560 a year, which represents more than a 40% reduction from prior years.

Furthermore, the comparison to IStation is not apples-to-apples, as Amplify's proposal offered more services, more data analysis, more assessment capability, and more materials than IStation. In particular, Amplify's offering provided to the State to date includes TRC, software that supports teacher administration of a reading running record. Teachers around the state rely on the software and book sets embedded in Amplify's software to assess students' reading levels to support day-to-day classroom instruction. In addition, unlike IStation, mCLASS provides more detailed reporting, with item level analysis of data collected in the DIBELS assessment. This data is much more valuable for teachers and literacy instructors to consider in targeting instruction, making the reporting more valuable. IStation does not report at the item level and therefore it is understandable that its data and reporting product might be less expensive.

In addition, there are hidden costs in any IStation implementation, such as headphones and microphones required for the computer-based administration. Purchase of software to replace these

gaps and hardware to support the new mode of administration will be borne by school districts, which did not budget for the change and may lack the resources to do so.

According to press reports, the State Board of Education intends to review this contract award at its July 11, 2019 meeting.

D. Amplify's Product

Amplify's product, mCLASS, is an assessment tool based on direct observations by a teacher of students performing authentic reading activities. In a series of reading tasks, the student reads directly to the teacher, and the teacher records student behavior into the mCLASS software. The teacher observes the student segmenting words into phonemes, sounding out words, reading words with automaticity, reading text (including illustrated children's books) with fluency and expression, and responding to comprehension questions. These activities are developed to model authentic reading behaviors, with detailed item-level data and related instructional recommendations provided to the teacher to drive the instructional experience for students.

mCLASS provides for streamlined data collection, emphasizing measures of the most important skills. The measures are administered in the manner that is most appropriate for the developmental stage of the child as well as the skills being assessed. Teachers have faith in the data collected via mCLASS, because it is the result of their direct interactions with authentic materials and a shared experience with the student, rather than only a number on a screen. This approach is at the heart of tailoring instruction to students learning the foundational skills. It also enables responsiveness to students who may need additional behavioral or socio-emotional support. This direct engagement with students' thinking during the assessment also provides more reliable results.

mCLASS's benefits have not only been seen in North Carolina schools, but nationwide. Amplify has partnered with schools across the country and has a demonstrated history of successful implementations. The recent experience of Colorado schools is one clear example. That state's department of education funded the use of mCLASS across the state for five years. At the beginning of the 2018-2019 school year, the department gave Colorado schools the choice between IStation's product and Amplify's product. 99% of school districts chose Amplify. Overall, Amplify provided tools to 149 Colorado school districts, covering 591 schools and 126,500 students. IStation was used in only seven schools, covering 1,200 students.

E. IStation's Product

IStation's assessment tool, ISIP, uses an entirely different model than Amplify. ISIP is purely software-based, so that the student does not read aloud and does not interact with the teacher. Instead, the student enters answers to questions, tapping or clicking on the device or selecting from a drop-down of multiple choices, on a computing device.

Comparing ISIP to the requirements of the RFP and the governing statutes, a number of deficiencies are apparent. ISIP does not meet the legal requirements for dyslexia screening. It does not provide sufficient data to allow determinations of short-term student progress, and it does not accurately and reliably determine which students will be reading on grade level by the end of the school year. It also fails to deliver the type of observational assessments that the RFP requires. IStation's computer-only driven product is not developmentally appropriate for the students in kindergarten through third grade that the RFP seeks to serve. ISIP also fails to assess certain measures required by the RFP. Under the requirements of the RFP, it is clear that IStation's product should have been disqualified or scored much lower than Amplify's product by the evaluation committee.

F. Evaluation Committee

Since the award announcement, Amplify has been concerned to learn that DPI's evaluation committee concluded as far back as December 2018 that IStation's product was inferior, did not meet the State's mandatory standards, and that Amplify should have been awarded this contract.

We understand that Dr. Amy Jablonski, who was employed by DPI as the Director of the Integrated Academic and Behavior Systems (IABS) Division, was a member of the evaluation committee. Though she recently left DPI, she publicly posted information about this procurement showing that the evaluation committee did not recommend IStation be awarded this contract, and shared those findings with DPI and the Superintendent during December 2018. Dr. Jablonski reports that the committee found that IStation's offerings were deficient in not "accurately determining risk in domains of reading," "screening for dyslexia (as required [by the RFP])", and in "provid[ing] needed tools for [specific learning disability] policy change." **(Exhibit F,** Social media posts by Dr. Jablonksi).

DPI has not yet provided the public records requested by Amplify regarding this procurement. But Dr. Jablonski's observations—and the additional information Amplify has gleaned from the evaluation process—are consistent with Amplify's knowledge of IStation's products in regard to North Carolina's specific requirements.

G. Protest Letter

Pursuant to the RFP and 09 NCAC 06B.1102, this protest letter is submitted within 15 days of the award announcement. Amplify asks that a protest meeting be scheduled by DPI as soon as feasible.

II.   Legal Standards For Procurements

North Carolina law requires public, open, and fair competition for government contracts. Competitive procurements allow the State to get the benefits of competition inherent in the marketplace: vendors competing on the quality of services offered, the cost of services, and in delivering the best value to the State. Thus, contract awards are subject to both administrative review and judicial review.

An agency award on a government contract must be reversed if it was awarded in a circumstance where the agency "(1) Exceeded its authority or jurisdiction; (2) Acted erroneously; (3) Failed to use proper procedure; (4) Acted arbitrarily or capriciously; or (5) Failed to act as required by law or rule," to the prejudice of the complaining party. N.C. Gen. Stat. § 150B-23(a).

In awarding a contract, an agency must also follow its own rules and regulations, including the rules contained in its own RFP. The failure to do so is reversible error. *See Humble Oil & Refining Co v. Bd. of Alderman*, 284 N.C. 458, 467, 202 S.E.2d 129, 135 (1974). In addition, where an agency has discretion, an agency's decision will be found to be arbitrary or capricious if it "indicate[s] a lack of fair and careful consideration" or "fail[s] to indicate any course of reasoning and the exercise of judgment." *Act–Up Triangle v. Commission for Health Servs.* 345 N.C. 699, 707, 483 S.E.2d 388, 393 (1997).

As this procurement was done under the delegation of the Department of Information Technology ("DIT"), DIT regulations set out the method for selecting a best value procurement, including the following:

- The agency's evaluation committee must evaluate bids based exclusively on the criteria stated in the RFP. 09 NCAC 06B.0302(1)(h).

- The committee must consider the "technical merit" of each bid, including the responsiveness of the bid to the specific purpose or objective of the RFP. 09 NCAC 06B.0302(1)(f)(ii).

- The agency must also consider factors including whether the vendor complies with industry standards, the vendor's past performance with the state, and the probability of the vendor providing the required service on time. 09 NCAC 06B.0302(1)(f)(iii).

- In evaluating the cost of a vendor's proposed solution, the committee should not merely consider the price of the product, but rather the "State's total cost of ownership." 09 NCAC 06B.0302(1)(f)(i).

When a procurement is "cancelled" and an agency proceeds into "negotiations," those negotiations remain subject to the confines of the original RFP. Thus, "negotiations should not materially alter the intent or scope of the original solicitation document." 09 NCAC 06B.0316(e).

A failure to meet mandatory RFP requirements cannot simply be fixed by cancelling a procurement. Furthermore, as to RFP requirements mandated by state law, these requirements apply regardless of any cancellation.

III.   The Award To IStation Does Not Meet The Requirements of North Carolina Law And Must Be Rescinded.

IStation's product fails to meet the specific requirements of the RFP and governing law, and therefore was ineligible for award. The winning vendor must "meet NCDPI's obligations under state laws," including under the RtA legislation. RFP, Section I ("Introduction"), at p.6 **Exhibit C**; *see* N.C. Gen. Stat. §§ 115C–83.1-115C-83.14. The solution must also comply with the 2017 statute regarding screening for dyslexia and other learning disabilities. *See* 2017 Sess. Laws 127 (**Exhibit A**).

IStation's ISIP product falls short of these requirements in several important ways. First, it does not distinguish with sufficient accuracy students that are at risk of not reading at grade level from those that are likely to be on track. Second, it does not meet the legal requirements for determining whether students are at risk of dyslexia. Third, ISIP's computer-based model does not meet the requirement that a screening tool be developmentally appropriate for students in kindergarten through third grade—this is the core problem in IStation's proposal that leads to other failings. Fourth, ISIP is insufficiently accurate at assessing or fails to assess certain required measures. Finally, ISIP's planned new voice-recording feature known as "ISIP-ORF" should not have been considered in determining whether IStation's product meets the requirements of the RFP and the legislation.

The difference between ISIP and mCLASS is not simply that mCLASS is more effective by some subjective measure. Rather, mCLASS meets the RFP and the General Assembly's mandates, while ISIP does not. DPI must therefore reverse its decision to award the contract to IStation.

A. IStation Does Not Adequately Predict Students Who Are At Risk of a Reading Difficulty

IStation's product is deficient and does not meet the requirements of the RFP and North Carolina law because it cannot demonstrate classification accuracy, a central component of the RFP and of any effective universal screening program. IStation therefore should have been disqualified from any award.

The RFP explains that classification accuracy "determines how well . . . the screener identif[ies] the two groups of students – those who are at risk for language impairment and those who are typically developing." RFP, Section C ("General Conditions for Proposals"), **Exhibit C,**

at 10. The RFP requires that bids demonstrate "how the proposed solution meets the requirements for a universal screener, including . . . adequate . . . classification accuracy." RFP, Attachment A, Spec. 4, **Exhibit C** at p. 25.

Classifying students into these groups with a high degree of accuracy is essential because the classifications are the basis for key instructional decisions. For example, at-risk students may be eligible for additional supports, such as additional intervention services in school or summer reading camps. *See, e.g.*, N.C. Gen. Stat. § 115C-83.6(a) (requiring local school districts to offer parents the opportunity to enroll students in reading camps if students' reading is below grade level). These interventions aim to ensure that every student is reading at grade level by the end of third grade. If that does not occur, students must generally be retained in third grade. N.C. Gen. Stat. § 115C-83.7(a), (b) (requiring retention in third grade for students not reading at grade level, unless a "good cause exemption" applies). Classification accuracy is therefore a crucial component of the RtA program. It is the gateway to intervention that will keep students on grade level. And classification accuracy ensures that resource-intensive measures—including in-school intervention, summer reading camps, and retention in third grade—are managed appropriately.

Amplify's proposal provided reliable, public data showing classification accuracy for all grades from K-3, but IStation has not. Classification accuracy is frequently measured by the "area under the curve" statistic ("AUC").[4] Commonly accepted standards require an AUC of 0.80 and higher. AUC figures for mCLASS are 0.90 for first grade, 0.88 for second grade, and 0.90 for third grade. In contrast, IStation has no known AUC figures for K-2.

Although neither vendor satisfies the AUC benchmark for kindergarten, there is a stark contrast between the demonstrated classification accuracy of the two products. IStation has provided no indication of the accuracy with which it predicts students' risk levels in K while Amplify has provided evidence of a degree of accuracy of prediction. The predictive validity evidence provided is additional evidence that the predictive capabilities of Amplify better meet the requirements of the RFP, Amplify's predictive validity correlation is 0.55 for kindergarten students, which shows moderate predictive validity. IStation's score of 0.39 shows that IStation is entirely ineffective at predicting future performance for kindergarten students. Because IStation's proposal cannot show classification accuracy and has limited predictive validity evidence, it should have been disqualified, and was not a product with the scope of the RFP's requirement eligible for negotiations.

We note that relatively low AUC and predictive validity scores are not surprising for kindergarten students. Because kindergarten students are often in the very early stages of their reading growth, classification accuracy will be low - the signals of reading risk are not yet strong,

---

[4] *See* Nat'l Ctr. on Intensive Intervention at Am. Institutes for Research, Academic Screening Tools Chart Rating Rubric,
intensiveintervention.org/sites/default/files/NCII_AScreening_RatingRubric_July2017.pdf

just like performance in the first mile of a marathon is not a stable indicator of how runners will perform in mile 24). However, the truly dreadful scores of IStation are perhaps explicable by the very concerns about developmental appropriateness that underpin this protest and the widespread teacher outrage. Kindergarten students simply don't perform reliably in computer based assessments. They need a human to help them understand the task, stay on task, and feel good about the task.

B. IStation Does Not Meet The Statutory Requirements For Screening For Dyslexia

IStation's product also fails to meet the statutory screening requirements for dyslexia. The RFP specifically required vendors to provide effective tools to allow teachers to screen for dyslexia. It required bids to "[d]escribe how the [product's] measures align with best practices and adequately and accurately identify indicators of risk for dyslexia in grades K-3." RFP, Attachment A, Spec. 8, **Exhibit C** at p. 25.

As the RFP recognizes, these capabilities are not only desirable but are also required by law. In 2017, the General Assembly passed a statute to ensure that all students with dyslexia "receive the necessary and appropriate screenings, assessments, and special education services to provide interventions for learning difficulties." 2017 N.C. Sess. Laws 127, § 1 (**Exhibit A**). The statute also requires local boards of education to ensure that the "diagnostic tools and screening instruments used for dyslexia . . . are age-appropriate and effective." *Id.* § 4.

The International Dyslexia Association ("IDA")—the authoritative voice on the latest and most reliable research and information on dyslexia and on necessary policy changes for dyslexia screening and intervention—provides guidelines for screening students in kindergarten through second grade for dyslexia risk. For each grade, the IDA provides a list of skills that an effective dyslexia-screening tool must measure[5]:

Grade	Skills to Measure
Kindergarten	- Phonological awareness including phoneme segmentation; - Blending, onset, and rhyme; - Rapid automatic naming, including letter naming fluency; - Letter sound association; and - Phonological memory, including non-word repetition.

---

[5] Int'l Dyslexia Ass'n, Universal Screening: K-2 Reading (2017), https://dyslexiaida.org/universal-screening-k-2-reading/.

First Grade	- Phoneme awareness, specifically phoneme segmentation, blending, and manipulation tasks; - Letter naming fluency; - Letter sound association; - Phonological memory, including nonword repetition; - Oral vocabulary; and - Word recognition fluency.
Second Grade	- Word identification, including real and nonsense words; - Oral reading fluency; and - Reading comprehension.

While mCLASS includes measures that directly comply with IDA screening guidelines—requiring students to directly perform the age-appropriate tasks to demonstrate their skills—ISIP does not. Indeed, mCLASS is an approved product by IDA, but IStation is not.

In particular, difficulty with phonological processing is a key indicator of dyslexia risk. IDA recommends assessing both phonological awareness and rapid naming as components of phonological processing. IStation only measures phonological awareness and in a way that requires students to match sounds in a multiple choice format rather than produce them. However, an accurate measure of phonological awareness not only requires that a student understand that words are made up of sounds, but also requires students to produce those sounds. IStation does not measure rapid naming, thus missing one of the skill measurements recommended by IDA. Research indicates that students who have difficulty with both rapid automatized naming and phonological awareness experience more reading difficulty and are more likely at-risk for dyslexia than students who have difficulty in phonological awareness alone (Wolf & Bowers, 2000; Ozernov-Palchik, Yu, Wang, & Gaab, 2016). By missing data on one of the skills, IStation's assessment risks missing identification of students at-risk for dyslexia, thus failing to comply with the legal requirement for an effective dyslexia screening instrument.

Because ISIP is non-responsive to the IDA's list of grade-appropriate reading constructs to measure, IStation's product does not meet the dyslexia-screening requirements established by the RFP and statute. ISIP's diagnostic tools for dyslexia are not "age-appropriate and effective," *see* 2017 N.C. Sess. Laws 127, § 4, because they do not "adequately and accurately identify indicators of risk for dyslexia," *see* RFP, **Exhibit C**, at p.25. Because the agency did not follow its own rules in selecting a vendor, the decision to award to IStation must be reversed. *See Humble Oil*, 284 N.C. at 467, 202 S.E.2d at 135.

The selection of ISIP will also negatively impact local school districts by failing to ensure age-appropriate screening instruments for dyslexia are made available. The General Assembly

required local school districts to "review the diagnostic tools and screening instruments used for dyslexia" before the 2017-2018 school year to "ensure they are age-appropriate and effective" and to "determine if additional diagnostic and screening tools are needed." 2017 N.C. Sess. Laws 127, § 4. At that time, all North Carolina school districts had access to mCLASS. ISIP's key diagnostic defects are not present in mCLASS. Thus, at the time the General Assembly required school districts to evaluate their diagnostic tools, those districts had access to an effective product.

Now, however, North Carolina school districts will be forced to procure another assessment for the purpose of screening for dyslexia, since they cannot rely on IStation and its tool, which is unfit for the task. Their forced transition to the inferior product, moreover—and the training that will require—will have to take place in an impossibly short period of time. Over a hundred year-round schools will start the new year in just a few weeks, and the rest will start in about 60 days. Thus, despite carrying out the evaluation required by the General Assembly in 2017, local boards of education will now be unable to meet the General Assembly's intent: to provide "the necessary and appropriate screenings, assessments, and special education services" to determine dyslexia risk, and provide appropriate interventions where the risk is high.

C. IStation's Computer-Based Only Assessment Model is Developmentally Inappropriate for All Students

The specific failings of IStation to show the accuracy of its assessment tool and to meet the statutory and RFP requirements for dyslexia testing are symptoms of the greater core problem with IStation's product. IStation's product relies on computers to the exclusion of teacher-student interaction, which simply is ineffective for the K-3 population. The computerized assessment model that makes ISIP inadequate for dyslexia screening also applies to ISIP's effectiveness in screening students generally for reading difficulties. In short, IStation's computerized model is not "developmentally appropriate" for the K-3 population in the absence of human interaction, and therefore IStation cannot meet the requirements of the RFP or the RtA legislation.

As discussed above, under North Carolina law, diagnostic reading assessments must be based on "developmentally appropriate practices." N.C. Gen. Stat. § 115C-83.6(b). The RFP therefore also requires that assessments meet that standard. RFP, Section I ("Introduction"), at p. 6 ("NCDPI is obligated to adopt and provide . . . developmentally appropriate assessments.") (**Exhibit C**). Bids were required to describe how their products use "developmentally appropriate practices to assess K-3 students." RFP, Attachment A, Spec. 9, **Exhibit C** at p. 26.

A robust body of research demonstrates what developmentally appropriate practices are for early reading skills. Assessments should be appropriate to the development of young children and should incorporate appropriate methods that include "observations of children, clinical interviews, collections of children's work samples, and their performance on authentic activities"

Procurement Manager
North Carolina Department of Public Instruction
June 24, 2019
Page 15

rather than artificial exercises.[6] This ties into the research—discussed above in the context of dyslexia—showing that many of the most important measures of early reading ability require observing a student's oral response to text.

To assess whether a solution offers "developmentally appropriate practices" for measuring key reading constructs—oral language, phonological and phonemic awareness, phonics, vocabulary, fluency, and comprehension— the *validity* of the solution's measurement of these constructs is critical.[7] Validity means the degree to which evidence supports a certain use of assessment scores and is widely recognized in the field of assessment as "the most fundamental consideration in developing tests and evaluating tests."[8] If an assessment fails to capture important aspects of a particular construct—because the assessment does not adequately sample some types of content, engage in certain psychological processes, or elicit response methods that are encompassed in the content—the assessment is said to have "construct underrepresentation."[9] The scores derived from that assessment are not a valid measurement of the construct.

The artificial constraints ISIP places on assessment activities give rise to serious questions as to ISIP's validity as a measurement of key reading constructs. One example may be found in the contrast between ISIP and mCLASS in their measurement of phonics, a construct specifically referenced in the RFP. *See* RFP, Attachment A, Specs. 1, 3, **Exhibit C**, p. 25. The National Reading Panel describes phonics as "the knowledge that letters of the alphabet represent phonemes [units of sound] and that these sounds are blended together to form written words."[10] A mastery of phonics allows a reader to "sound out words they haven't seen before, without first having to memorize them." To directly measure phonics, an assessment must give students the opportunity to blend letter sounds into words.[11] This is exactly how mCLASS measures phonics ability. ISIP,

---

[6] Nat'l Ass'n for the Educ. of Young Children, at 22.

[7] Nat'l Ass'n for the Educ. of Young Children, *s*at 22 ("Sound assessment of young children is challenging because they develop and learn in ways that are characteristically uneven and embedded within the specific cultural and linguistic contexts in which they live. For example, sound assessment takes into consideration such factors as a child's facility in English and stage of linguistic development in the home language. Assessment that is not reliable or valid, or that is used to label, track, or otherwise harm young children, is not developmentally appropriate practice.")

[8] AERA Standards.

[9] AERA Standards.

[10]Nat'l   Institute   of   Child   Health   &   Human   Dev.,   Nat'l   Reading   Panel, https://www.nichd.nih.gov/research/supported/nrp.

[11] The North Carolina English Language Arts Standard Course of Study explicitly calls for students to demonstrate skill in both blending and segmenting words at the onset-rime and phoneme level. mCLASS includes the Phoneme Segmentation assessment that asks students to segment words at the phoneme level and allows for observation of whether students are able to segment at the onset-

on the other hand, does not involve any oral production of letter sounds or the blending of sounds into words. Instead, its measure of phonics is whether a student can click on the correct reading from options on the screen. The validity of this task for assessing phonics is doubtful.

ISIP suffers similar validity problems for other constructs listed in the RFP. For example, phonological and phonemic awareness are ways to measure whether young children understand the sounds in words, even before they learn to read print. *See* RFP, Attachment A, Specs. 1, 3, **Exhibit C**, p. 25. Again, choosing between options on a screen, without being asked to produce sounds, severely restricts the assessment of this construct.

Oral fluency is another key construct. *See* RFP, Attachment A, Specs. 1, 3, **Exhibit C**, p. 25. It has long been recognized that oral fluency is best observed by listening to a student read aloud.[12] Requiring students to instead fill gaps in sentences by choosing from words on a screen simply cannot replicate reading aloud.

IStation's product may be able to measure certain limited aspects of reading ability, but in the absence of in-person observation, its measurements will necessarily be incomplete. For example, ISIP does not allow students to demonstrate that they can produce sounds as a component of phonological or phonemic awareness. Instead, students must simply select the correct sound from a small set of options on the screen. IStation's constrained measures are not developmentally appropriate.

Moreover, effective assessment must include direct observation. Reading assessments do not occur in a vacuum. Teachers are tasked with helping students improve, especially those at high risk of reading difficulties, between the assessments. Thus, teachers and students benefit when assessments help determine not just *whether* a student may have difficulties, but also why. Some students may score poorly on a reading assessment, for example, because they are distracted by external activities or objects. Students will also approach the text differently. Some may move quickly past difficult words without engaging for any length of time; others may linger on difficult words and be unable to complete the passage. These situations call for very different responses, whether that means helping a student to focus appropriately on the assessment or using tailored approaches to teaching between assessments. For example, the "three-second rule" in the standardized administration of mCLASS provides that a teacher assessing oral fluency should prompt a student to advance past a word she cannot recognize after three seconds.

It is no coincidence that the RFP asks vendors to describe their products' "observation-based practices." The RFP contemplates that observation will be a feature of a model that reliably

---

rime or phoneme level, including full segmentation. The IStation assessment includes phoneme blending and rhyming, skills that are not as advanced as the NC Standards call for.

[12] *See, e.g.,* Jerry Zutell & Timothy V. Rasinski, Training Teachers to Attend to Their Students' Oral Reading Fluency, 30 Theory Into Practice 211 (1991).

assess the key indicators of reading difficulty. *See* RFP, Attachment A, Specs. 1-2, **Exhibit C**, at 25.

While it is true that the RtA legislation states that diagnostic assessments "may be administered by computer or other electronic device," *see* N.C. Gen. Stat. § 115C-83.6(b), this language does not sweep aside IStation's deficiencies. Amplify does not take the position that no effective assessment can have a computer-based element. Indeed, Amplify's advanced software is an important component of its own solution. Teachers enter data electronically even during in-person assessments, which saves time and facilitates progress tracking. Amplify even offers a rigorous online assessment option for assessing some of the same skills as can be assessed in person. Crucially, however, this is to be used by students who are already reading on track and in grades 2 and higher. The net effect of having both types of assessment available is that teachers can tie both online and observational results to personalized instruction for all students in their classrooms, regardless of individual skill levels. Thus, as the statute recognizes, partially computerized solutions may well balance the needs to optimize teacher time, gather advanced metrics, and give students the appropriate in-person instruction time. The touchstone for any solution, however—as the statutory language makes clear—is that it uses "developmentally appropriate practices." *Id.* IStation's computer-*only* product does not meet this prerequisite.

Therefore, because IStation's offering is not developmentally appropriate under state law and the RFP, DPI should have disqualified IStation's proposal, or should have rated Amplify's product as far superior. Indeed, we understand that one of the reasons the evaluation committee recommended against award to IStation, and for award to Amplify, is because IStation's program is not developmentally appropriate.

D. IStation Fails to Assess Certain Required Measures

IStation also fails to meet the RFP's requirement for oral language measure, defined as expressive and receptive. *See* RFP, Attachment A, Spec. #1a, **Exhibit C**, p. 25. IStation's Oral Language measure requires students to listen to a sentence and identify the picture that best illustrates the orally read sentence's meaning. Another task in that measure requires the student to choose a word that best completes the sentence or story read aloud to the student. These tasks only satisfy the "receptive" prong of the requirement. IStation's measure fails to provide an opportunity for students to demonstrate expressive language. In either task, the student is only demonstrating listening comprehension, not expressive language.

By contrast, in mCLASS, a teacher reads a sentence aloud and student is asked to repeat the sentence back to the teacher. The teacher captures the errors and miscues of student response, with the error patterns then analyzed to inform areas where the student may need further development in the student's oral language. Thus mCLASS is assessing both receptive and expressive oral language.

IStation's failure to satisfy the requirement for an expressive and receptive oral language assessment is another area where IStation's product is not responsive to the RFP, and a reason why IStation's proposal should have been disqualified.

IStation's product also failed to meet the RFP requirement to measure reading rate and accuracy in several of the key domains of reading for students, at the very least in the areas of phonological and phonemic awareness and phonics. The RFP recognizes that a complete picture of a student's performance on any given measure must consider a student's reading rate as well as accuracy. The RFP therefore required vendors to describe "how the proposed solution measures accuracy and rate" for students' oral language, phonological and phonemic awareness, phonics, vocabulary, fluency, and comprehension. *See* RFP, Attachment A, Spec. 2, **Exhibit C**, at 25.

These terms have well-defined technical meanings: accuracy is the percent of correct items out of all items attempted, and rate is number of correct items per unit of time.

In all relevant mCLASS measures, rate and accuracy are a key element of the scoring. ISIP, on the other hand, seems unlikely to be able capture a student's rate and accuracy. A computer-based assessment cannot deliver an unmediated accuracy and rate measure, because of the time required for the student to enter responses into the computer - *e.g.* a student may perform the reading task swiftly but enter answers slowly (which would lead to a rate calculation that is artificially low.) Similarly, on a computer-based assessment it is not always possible to know exactly which elements of a reading passage or which phonetic or phonological elements of a word a student has read successfully en route to answering the eventual multiple choice question. To be concrete: in ISIP's reading fluency measure, the students read five or so words on the way to the drop down menu where they have to select the sixth word. One can tell if the student selected the right sixth word, but one can't really tell which of the five previous words were read accurately. Did they fail to read all five words, which would be an accuracy of 0%, or did they read three words correctly, which would still be an accuracy of 60%? ISIP cannot capture this information.

IStation may argue that their measures correlate with accuracy and rate, but the RFP asks for accuracy and rate, not measures that correlate with accuracy and rate. In this sense, the ISIP computer based assessment is not compliant.

It may be possible that ISIP includes some measures that correlate with reading rates, but in fact are not reading rates. This is another area in which IStation's product is not responsive to the RFP, and a reason why IStation's proposal should have disqualified or rated as inferior.

E. IStation's Product Is Ineffective at Measuring a Student's Progress

It is our understanding that IStation's measurement of student progress is cumbersome and lacks increments required by the State. Effective universal screening cannot rely on a static

measure of performance at any point in time; it must also measure progress at frequent intervals.[13] Thus, the RFP explicitly stated that the solution "must assess student progress," RFP, Section I ("Introduction"), **Exhibit C** at p. 6, and provided a list of criteria a vendor had to meet to provide sufficient progress monitoring, *id.* Attachment A, Spec. 6, at p.25. These included the requirements that progress monitoring be "brief," "repeatable," and "sensitive to improvement over time, including short-term change." *Id.*

Consistent with these criteria, Amplify's mCLASS program allows for testing on a weekly basis and is sensitive to weekly growth. Moreover, a teacher using mCLASS can complete the testing for each measure, to provide comparisons to prior weeks, in one minute. In contrast, ISIP's testing for progress monitoring takes approximately 30 minutes. Even then, a teacher may not evaluate growth through ISIP until the student has completed *three* consecutive *months* of assessments. In comparison, Amplify's mCLASS DIBELS assessment provides exactly the sensitivity required by the RFP, where growth can be evaluated in as little as two weeks.

IStation cannot satisfy the RFP's requirements because it cannot show that ISIP is sensitive to "short-term change." RFP, Attachment A, Spec. 6, **Exhibit C**, at p.25. IStation's more cumbersome mechanism for monitoring progress also does not satisfy the requirements that the testing be "brief" and "repeatable," and decreases the possibility of teachers being able to effectively oversee reading improvement. For this reason, as well, the contract award should be reversed.

F. IStation's Purported New Feature Should Not Have Been a Factor in the Contract Award

To the extent DPI considered a new feature called "ISIP-ORF" that IStation plans to roll out for ISIP, that consideration was improper. IStation apparently plans to introduce this measure at some point in the future. ISIP-ORF is IStation's new oral fluency measure that uses speech recording and speech recognition technology. IStation's press release regarding the contract award appears to refer to alleged benefits of ISIP-ORF. The press release states: "The data collection stage of the program allows students in grades one through three to record a reading passage by logging in to the program and using a microphone headset." (IStation Press Release, May 9, 2019, **Exhibit G**.) For several reasons, however, DPI should not have relied on the availability of ISIP-ORF in awarding the contract.

---

[13] Douglas D. Dexter & Charles Hughes, *Progress Monitoring Within a Response-to-Intervention Model*, RTI Action Network, http://www.rtinetwork.org/learn/research/progress-monitoring-within-a-rti-model ("Progress should be monitored frequently, at least monthly, but ideally weekly or biweekly."); U.S. Dep't of Educ., Nat'l Ctr. for Educ. Evaluation & Regional Assistance, *Practice Guide: Assisting Students Struggling with Reading: Response to Intervention (RtI) and Muli-Tier Intervention in the Primary Grades*, 24-26 (2009) (Recommendation 4, that progress should be monitored at least once a month).

First, Amplify believes ISIP-ORF is still in the research phase. There has been no public data on the functioning or effectiveness of the new feature, which should be an essential prerequisite to using the feature in schools. If ISIP-ORF is not yet functioning, its supposed benefits will not be available to teachers or students and those factors should not have been factored into the evaluation of the bids. Moreover, even if IStation has provided DPI a confirmed rollout date for the new feature—and even if IStation can meet that deadline—the rollout will add significant additional implementation and training time.

Second, the use of ISIP-ORF undermines one of IStation's supposed benefits: giving time back to teachers. The program records a student's reading via a microphone connected to the computer. The teacher must then go back and listen to the recording. In contrast, with mCLASS, reading and assessment are combined into a one-step component. IStation's new "feature" appears to have sacrificed direct interaction with students, in the name of reducing teacher instructional time, merely to replace it with more teacher time away from students. Though IStation may hope to solve this problem by rolling out automated scoring, Amplify is aware of no evidence that ISIP-ORF's automated scoring is reliable. Indeed, previous research on automated ORF scoring has shown that this approach is flawed, especially for students with accents or other speech issues.

Third, though ISIP-ORF requires more teacher time than ISIP's existing model, the new feature fails to deliver the benefits of in-person instruction. This defect will remain regardless of whether a student's recording is scored manually or automatically. As one example, a valid measure of oral reading fluency requires a teacher's direct supervision. One important principle of oral fluency assessment, as mentioned above, is the "three-second rule": a teacher must prompt a student to move along if she spends more than three seconds on a single word. Another important principle is that a student's repetition of a line or sentence does not affect her score. We are not aware of ISIP-ORF accounting for these principles in its current state. Thus, IStation's product still fails to meet the standard for developmental appropriateness that the RFP requires.

Fourth, the mCLASS solution includes tools that allow for specific error analysis of student reading skills and behaviors that are directly observed by the teacher. The system includes the Item Level Advisor which uses the logged observations from student responses to DIBELS Next benchmark and progress monitoring assessments to identify specific response patterns in a child's results (such as a child sounding out nonsense words sound-by-sound rather than for each as an entire word) and to recommend reinforcement activities tied to the specific performance pattern. In addition, teachers can conduct an MSV analysis on student miscues from TRC to determine whether they happened based on Meaning cues (e.g., substituting "kitty" for "cat"), Structural cues (e.g., substituting a noun for another noun), or Visual cues (e.g., substituting a word like "cat" with a similarly spelled word like "cap"). While, ISIP ORF claims to offer "computer-based resources and formative data to identify both when a student makes a reading error and what kind of error— be it self-correcting, visual, meaning or syntax, it is unclear what level of analysis is provided by

ISIP ORF, and published efforts to do this sort of analysis automatically have indicated serious problems with the approach.

For all these reasons, the new ISIP-ORF feature should not have been a factor in selecting an awardee.

IV. DPI Must Provide Justifications for Its Decision to Award the Contract to IStation

DIT rules require that the reasoning behind a contract award, along with evidence supporting that reasoning, must be made available to the public. *See* 09 NCAC 06B.1402(c); *id.* 06B.1402(b). To date, DPI has provided no such documentation, and thus has shown no reasoning to justify an award to IStation. Any decision to cancel a procurement must also be documented with reasoning, but DPI has not produced any information to justify the cancellation. *See* DIT Procurement Office Manual (July 2017) ("Justification and documentation regarding the cancellation must be included in the Procurement File."). On the current record, there has been no basis offered by DPI to show that its award meets the requirements of North Carolina law.

Amplify has made public records requests to DPI, and reserves the right to supplement this letter if DPI produces public records regarding the basis for RFP cancellation and ultimate award.

V. The Implementation of IStation Should Be Stayed, and Amplify's Contract Should Be Extended, While the Protest Is Pending

The award announcement on June 7 allows little time for schools to implement IStation's product and train teachers on how to effectively use it. There are approximately 150 year-round schools in the State with start dates in July. Amplify has heard from several school districts that they are concerned about their ability to adequately assess students using the new software and mode of administration, given that the districts were not given time to plan for the implementation. In Amplify's experience, a statewide implementation of the magnitude in North Carolina takes many months, with careful planning for training and onboarding, so that educators are comfortable and proficient at leveraging the program to gain meaningful data necessary for targeting instruction, monitoring their progress and identifying students who need additional support. This is especially so for a program that radically changes the mode of assessing literacy skills and provides less information for teachers, who are used to getting item-level data and reading level data to inform day-to-day instruction. A new implementation also requires time for district and school leaders to establish the systems of support for the school system as a whole as the assessment data supports many efforts across the school system. Leaders need to evaluate processes related to the creation and monitoring of IEPs for students with special education needs, so all of the IEPs that have goals that are related to mCLASS assessments results would need to be updated. In addition, LEAs and schools will need to revisit policies surrounding their MTSS (Multi-Tiered System of Supports) procedures, the establishment of individual reading plans for students, ingestion of data into their local data systems, teacher evaluations, parent

communications, and specialist and administrator training, to name a few. LEAs and schools also need time to set policies around how the new data and results will be used in the daily instructional processes, planning, and decision-making around students, *i.e.*, which specific interventions or other services are needed for struggling students who need support to get on-track.

Amplify requests that the contract award to IStation be stayed during this protest. In addition to the implementation concerns noted above, press reports indicate that the State Board of Education intends to review the IStation contract as its July meeting. A stay of implementation is not a rescinding of an award. It is a decision to keep the status quo in place so that an agency can make a reasoned decision with appropriate consideration.

VI. Conclusion

For the reasons set forth above, Amplify submits that the award to IStation was in violation of North Carolina law, the RFP, and DIT rules, and was arbitrary and capricious. Therefore, the award to IStation should be rescinded.

Furthermore, Amplify requests that a protest meeting be scheduled as soon as possible, and that performance of IStation's newly-awarded contract be suspended until this protest is resolved. Because school is starting very soon and the State Board of Education intends to further review the award, forcing a rapid implementation of a new product while a protest is pending (and while most teachers are not at work and not available to be trained) would be ill-advised.

Thank you for your attention to this matter. We look forward to meeting with you to discuss this procurement.

Very truly yours,

SMITH, ANDERSON, BLOUNT, DORSETT,
MITCHELL & JERNIGAN, L.L.P.
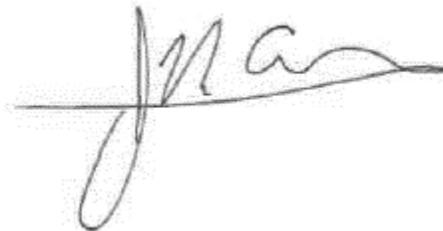
By: J. Mitchell Armbruster

JMA/efr
Enclosure (exhibits)

cc: Jonathan Sink, General Counsel, DPI   (via email: Jonathan.Sink@dpi.nc.gov)