

Introduction

Interim or interim benchmark assessments are a recent phenomenon in American education. Until about a decade ago, large-scale assessment occurred once per year in states and districts that chose to administer such assessments. Two federal reauthorizations of the Elementary and Secondary Education Act (the 1994 Improving America's School's Act and the 2001 No Child Left Behind Act) changed this. By requiring states to set academic content standards for all of its schools in English Language Arts/Reading, Mathematics and Science, to develop assessments aligned to these content standards, and use aligned academic performance standards to gauge the performance of students (and ultimately, that of America's schools), it made it possible to judge whether each school was making "adequate yearly progress" (meeting or exceeding the ever-increasing bar set for school performance under a formula determined by federal law). Certainly, the stakes for performance by states and schools have increased dramatically.

*"Interim or interim
benchmark assessments
are a recent
phenomenon in
American education."*

In response to these increased stakes, schools began to realize that waiting to find out how improvement activities were working until the annual statewide assessment was a risky proposition. Instead, they began assessing students one or more times before the annual statewide assessment - with enough time to intervene, so as to be able to show improvements in student performance. Thus, "interim" assessments were born. Adding impetus to this movement were the results from large urban schools districts that seemed to show that such benchmark assessments, combined with other interventions, led to noteworthy and significant improvements in student performance. While once the domain of such urban districts, these interim assessments are now part of a more balanced approach to assessment found in many schools of all types.

As schools have begun to use such assessments, several different variations of these assessments have been developed and used. Because these variations differ both in purpose and fundamental design, it is vital to consider these, especially in instances where new programs of such assessments are being considered. Hence, the purpose of this paper is to describe several different ways that interim or interim benchmark assessments can be developed, along with the purpose(s) for such different designs.

Introduction to IRA Design

The "traditional" design for interim benchmark assessments has been the

one in which students are assessed each quarter of the school year to gauge their likelihood of scoring at the proficient level or above on the statewide assessments. The purpose of this sort of IBA is predictive; presumably, students not scoring at a level that is likely to result in a proficient or above score on the annual statewide assessment will be afforded additional instructional opportunities designed to improve their achievement - and scores on the annual assessment. Of course, the downside of this approach may be for educators to identify and work with only those students whose predicted scores are close to the passing level (the so-called “bubble kids”) and ignore those that are not at all close to scoring at this level.

However, IBAs need not (and have not) stuck to just this one model. There are other purposes for such assessments, and these purposes result in somewhat different assessment designs. These assessment designs have implications for assessment development and assessment administration. Thus, all of these models should be considered when IBAs are to be developed, so that the IBA design matches the intended purpose(s) for the assessment.

Purposes and Designs

As mentioned above, there are several different IBA models. The purposes for these different designs, as well as the designs themselves, are described below. The purpose of this section is to describe different uses for IBAs so that clarity in design is used to assure that the intended purposes can be met by the IBA. This descriptive material is not intended to convey that one approach to IBAs is better than another, or is to be preferred, just that they are different.

Predictive IBA - As described above, this type of interim benchmark assessment is the traditional approach to these assessments. In this model, three or more versions of an assessment designed to predict overall performance on the statewide assessment given at the end or the beginning of the school year, are developed and validated through a study to determine how well performance on each IBA instrument predicts overall statewide assessment performance. To be maximally predictive, these IBAs should cover the same academic content standards as those included on the statewide assessment program. Of course, the number of assessment items may be fewer in number total, so that it is feasible to administer the assessment in one or two class periods. These assessments may be given quarterly or at the conclusion of each marking period.

Progress Monitoring IBA - The goal of a progress monitoring IBA is to be able to show the progress of students in learning the content covered by the academic content standards assessed on the statewide assessment over an entire school year. What is reported is the student growth in learning across three or more assessments conducted during the school year. While ultimately educators will be concerned whether the level of performance of students at the conclusion of the IBAs is consistent with passing the statewide assessment program, this is not the ultimate issue; rather it is the “growth” in performance across the school year that is of concern. In this design, each IBA measures the same academic content standards with related items. In this manner, the “progress” of the student can be readily demonstrated through their achievement of more and more of the standards throughout the school year.

“This descriptive material is not intended to convey that one approach to IBAs is better than another, or is to be preferred, just that they are different.”

“All MAEIA documents are available at: mi-arts.wikispaces.com.”

Pacing IBA - In a pacing IBA, each examination used covers the instructional material contained in one portion of the year’s instruction, such as a marking period or a semester. One way this is done is by slating each IBA to cover a particular portion of the school year. The presumption of this model is that instructional outcomes learned in each instructional period will add up to overall satisfactory performance on the statewide assessment. A key goal of this model is to assure comparable opportunity to learn for all students - that all teachers instructing students at the same grade level and content area or in the same course will provide consistent instruction on the selected instructional goals. Thus, while educators will attend to whether students are achieving the goals on which they were instructed, administrators may be attending to the extent to which consistent instruction was provided to students, as evidenced by differences in their performance across similar classrooms or schools.

Instructional IBA - Another type of IBA is assembled to measure the outcomes of each instructional unit across a school year or a course at the secondary level. At the secondary level, these instructional IBAs may be in addition to or in place of more traditional end-of-course examinations. To use this type of IBA, it is necessary for educators to group the academic content standards for a course or a school year into units around which instruction is to be built. The number and complexity of the content standards, as well as the time needed for adequate instruction on the unit, will dictate how long the unit will last and when the instructional IBA will be administered. Each IBA will be administered at the conclusion of the instructional unit and is often used to grade the students on their achievement of the content included in the unit of instruction.

Informative IBA - This type of IBA consists of many small testlets or even test items that are administered often - sometimes, on a daily basis or several times per week. As in true formative assessment, the goal of using these IBA “mini-assessments” is to determine whether the instruction provided by a teacher each day has been effective for all students to learn, and if not, to assist the teacher to gauge where there are issues in student learning so that additional learning opportunities can immediately be provided to students. Thus, this type of IBA is not summative in nature such as those listed above, but instead, is designed to be used by teachers for immediate instructional interventions with students. Such formative IBAs, used daily, may be followed by one or more of the other types of IBAs upon the conclusion of the instructional unit or a particular amount of time (e.g., a marking period).

IBA Assessment Designs and Content

Based upon the types of IBAs described above, there are different assessment designs and typical content for each of these IBAs. These are described below and explained further.

Predictive - This type of assessment will be a mini-replication of the annual summative statewide assessment. One version of this design will be selection of the subset of content standards and assessment items that

best predicts overall statewide assessment performance. Another is to include the same set of content standards as on the statewide assessment.

Content - The content standards that are selected for this type of assessment may be based on the subset of assessment items needed to predict overall statewide assessment performance, or this assessment may appear to be a replica of the same content and item formats included on the statewide assessment.

Progress Monitoring - This type of assessment will be a mini-replication of the annual summative statewide assessment. The assessment might be a one-for-one replication of the statewide assessment. It may also include even more content standards (those leading to accomplishment of standards assessed on the statewide assessment).

Content - Each form of the IBA assessment used throughout the school year will cover the same set of content standards. Typically, about three forms will be used annually. All of the content standards assessed on the statewide assessment will be included in these IBAs. Additional standards that lead to accomplishment of state outcomes may also be included.

Pacing - The assessment will cover the content standards found in each period of instruction (e.g., a single marking period). These may include the content standards assessed in the statewide assessment taught during the period of time, plus enabling skills that lead to the accomplishment of these skills.

Content - Each form of the IBA assessment used throughout the school year will cover the same set of content standards. Typically, three to four forms will be used annually, although additional (or fewer) forms might also be used. The key content standards addressed in each period of instruction will be addressed in these IBAs. There may be other standards addressed during instruction (e.g., enabling skills) that are not included.

Instructional - This assessment will use a variety of assessment measures to assess key instructional outcomes that should be achieved in instructional unit. A variety of assessment methods may be used on the statewide assessment, such constructed-response and performance measures.

Content - Each form of the IBA assessment used throughout the school year will cover different content standards. Typically, three to four forms will be used annually. The key content standards addressed in this single unit of instruction will be addressed in these IBAs. These may include key outcomes and the enabling skills leading up to them in key learning progressions.

Informative - This system of assessment items will be comprised of a series of single test items (especially if they are performance assessments or constructed-response items) or mini-tests (two or three test items). Each item or mini-test should address a single content standard or a series of highly related content standards in a single learning progression. These items are designed to be used daily or several times per week, and are not to be confused with longer sets of items that address one or more of the IBA types described above.

“Based upon the types of IBAs described previously, there are different assessment designs and typical content for each of these IBAs.”

Content - Each form of the IBA assessment used throughout the school year will cover different content standards. There may be as many as 8 to 12 (or more) forms created for a course taught over a full school year. To increase flexibility in this type of IBA, testlets measuring individual standards might be created so that as different teachers teach the content standards in different combinations and sequences, they can group the corresponding testlets into end-of-unit IBA exams. These IBAs are designed to be used frequently, within the context of daily instruction. For example, these items might be used as a “question out the door” at the conclusion of each class period or instruction in a content area. These assessment items will be used and reported individually or in small clusters, and will be designed to provide immediate, actionable feedback to classroom teachers. Thus, they are more in the form of individual test items, not tests.

Strengths and Challenges of Each type of IBA

“Users will want to be mindful of both the strengths of each approach and the challenges of using them when selecting a type of IBA for use.”

Each of the IBAs listed above have purposes and uses, as noted. However, each also has some drawbacks. Users will want to be mindful of both the strengths of each approach, and the challenges in using them, when selecting a type of IBA for use. While there may be some ways of ameliorating the negative impacts of each approach to IBA, not all such challenges can be adequately addressed with just a modified IBA design. Some such challenges may require additional assessment approaches, such as formal summative assessments at one extreme and learning to use true formative-assessment strategies connected to daily instruction at the other. The strengths and challenges of each type of IBA are given below:

Predictive - As mentioned earlier, the goal of this type of IBA is to predict how students will do on the annual statewide assessment so that students predicted to do poorly (score below the proficient level) can be identified and receive extra instructional assistance.

Advantages - This approach should permit educators to identify students in need of assistance before they do poorly on the statewide assessment. This should move more students to learn the material that forms the basis of the statewide assessment and thus, both the student and the school will benefit from higher achievement.

Challenges - The challenge with this type of IBA assessment is that students who are not close to scoring at the proficient level may be ignored by educators since their chances of scoring at the proficient levels are slim. This may lead to the students who need the greatest level of support not receiving such assistance (and even less assistance than they otherwise might have received). Also, students who are predicted to score at or above the proficient levels may not receive as much support in their quest to continue to learn, since more attention is directed to the students just below proficiency.

Progress Monitoring - The goal of a progress monitoring IBA is to be able to show the growth in learning of students on the content covered by the academic content standards assessed on the statewide assessment over a school year.

Advantages - The improvement in performance across the school year is the focus of these assessments. Examining the changes in learning is important,

and good quality IBAs can assist with this.

Challenges - One of the key challenges in this approach to assessment is to be able report growth in a reliable manner. Changes scores on assessments are notoriously unreliable. We may (or may not) see changes that are due more the chance factors than true changes in student learning. Another challenge is to determine how much growth is enough for us to be satisfied that students are on target to improve sufficiently. Growth may not be consistent across the school year, either, so some students may evidence “growth spurts” while the growth of other students could be characterized as “slow but steady.” This is not the sort of metric that is often computed, so there is little or no prior information on which to base judgments about adequacy.

Pacing - In a pacing IBA, each examination used covers the instructional material contained in one portion of the year’s instruction. This might be done by designing four different IBAs, each covering a quarter of the school year. In order to create these four instruments, someone at the district or state level will need to determine the academic content standards that all teachers must address and that students should learn in each instructional period.

Advantages - The advantage of this approach is that it will standardize what teachers at the same grade or teaching the same course are to address, with considerable standardization of the order in which they address the academic content.

Disadvantages - Of course, this standardization is also a disadvantage of this model of IBA as well, since teachers often report (rightfully or not) that such standardized approaches are not suitable for students because students do not all learn at the same pace. Some students are struggling (since there is not time to pause to help these students achieve the content they have not achieved) while other students who are doing well could move at a faster pace and begin work on standards covered in the next quarter. The pace of instruction may not permit teachers to have the opportunity to enrich instruction for these students.

Instructional - The instructional IBA is designed to measure the outcomes of each instructional unit across a school year or a course at the secondary level. To use this type of IBA, it is necessary to group the academic content standards into units around which instruction is built.

Advantages - The major advantage to this type of IBA is that by the state or district created end- of-instructional unit exams, the quality of the assessments used to grade students at the end of such units could be vastly improved. These IBAs can also assure that all students are given comparable opportunities to learn. These IBAs could be created in ways to give educators some flexibility in the order of instruction, without losing the advantages of standardization. One way to do this is to develop a mini-assessment or testlet for each content standard, rather than a priori grouping content standards into instructional units. Each instructor could then select the combination of content standards he or she feels is important to be taught together (and sequence these units of instruction as he or she feels best), while at the conclusion of each unit of instruction (when it is time to assess student mastery of the instructional material in that unit), the instructor could use the corresponding testlets for the standards in that unit

“Advantages and disadvantages of each type of IBA design.”

“The goal of this paper is to try to distinguish different types of IBAs and how they might be designed and implemented so as to show that there is not just one way or one purpose for such assessments.”

to construct an assessment to measure students’ accomplishments. This has the advantage of standardizing the testlets (which would be validated before initial use) without the disadvantage of forcing instructors to use the same combination of standards in the same order.

Challenges - One of the challenges in building such a system of IBAs is to specify which academic content standards “go together” into each instructional unit. This can be a serious educational if not political dilemma for policy makers and educators, given the country’s reluctance for states and national entities to specify curricula and instructional designs for local educators’ use, and local schools reluctance to accept and use such standard curricula if offered. If such standardization is done locally, this would involve considerable duplication across districts within a state and across states.

Informative - This type of IBA consists of many small testlets or even test items; these are administered often - sometimes, on a daily basis or several times per week - by classroom teachers, with data available immediately to the instructor. As in true formative assessment, the goal of using these assessments frequently is to determine whether the instruction provided by a teacher has been effective for all students, and if not, to assist the teacher to gauge where there are issues in student learning so additional learning opportunities can immediately be provided.

Advantages - The major advantage to this approach to assessment is that it will encourage teachers to determine students’ accomplishment of instruction provided daily and permit teachers to adjust instruction in an on-going manner so as to assist most if not all students to achieve the intended content. The use of items that have been validated will help assure that students who do poorly on these items do so because of lack of achievement, not poor quality items or tasks that are confusing to them.

Challenges - The use of individual test items to measure students can be unreliable. The unique content of a single item, used alone to measure students, may lead to a false negative judgment of student accomplishment (we falsely conclude students did not get what we taught them today) or a false positive (where we erroneously conclude students had learned the material, but have not fully comprehended the concepts we taught them). In addition, individuals without much assessment experience may not be able to select assessment items appropriate for the intended assessment purposes; items selected may not align well with content standards being measured and/or may not be adequate assessment items.

Summary

The typology and descriptions of interim benchmark assessments are provided in this paper to distinguish several different types of such assessments. The goal of this paper is to try to distinguish different types of IBAs and how they might be designed and implemented so as to show that there is not just one way or one purpose for such assessments. It is also hoped that this typology can be used when research on the effectiveness of IBAs is carried out - to help distinguish that different IBA designs can have different uses and impacts. By describing these different types of IBAs as well as their content, the goal is to help differentiate different approaches to such assessments and how each type might be used in constructive student and school improvement activities.