HOW MIGHT THE USE OF TECHNOLOGY IN FORMATIVE ASSESSMENT SUPPORT CHANGES IN MATHEMATICS TEACHING?

SHAI OLSHER, MICHAL YERUSHALMY, DANIEL CHAZAN

Reform efforts in mathematics education since the late 1980s have sought to change how mathematics is done in classrooms—to reduce the focus on technique and to encourage creativity. For example, in the US, the NCTM Standards movement sought to encourage classrooms where students present multiple solutions to open-ended tasks and then engage in discussion to justify and critique these solutions (NCTM, 1991). Similarly, the US Common Core State Standards Initiative (National Governors Association Center for Best Practices/Council of Chief State School Officers, 2010) includes the Standards for Mathematical Practice that describe classrooms in which, for example, students use appropriate tools strategically, persevere to complete demanding tasks, construct viable arguments and critique the reasoning of others, and attend to precision. Still, teachers face challenges when trying to conduct lessons that productively build on student-designed solution strategies (e.g., Ball, 2001). In this article, we focus on formative assessment, conceptualized as: “all those activities undertaken by teachers, and/or by their students, which provide information to be used as feedback to modify the teaching and learning activities in which they are engaged” (Black & Wiliam, 1998, pp. 7-8). Specifically, we explore how technological supports for classroom formative assessment (Stacey & Wiliam, 2013) might play a role in attempts to reform mathematics instruction, particularly if they provide real-time feedback on complex student performances that can be immediately used in the service of instruction.

To do so, we use the construct of instructional exchanges (Chazan, Herbst & Clark, 2016) that:

- are marketplaces managed by teachers; teachers recognize in the midst of students’ mathematical activity those actions taken by students that “trade” as indicators of the acquisition of the knowledge that teachers are supposed to teach. In that sense, these student actions have become academic “gold”. (Buchbinder, Chazan & Fleming, 2015, p. 2).

A challenge for teachers teaching in the institutional context of contemporary compulsory public schooling is to manage such exchanges given the quantity of student work produced by students during a school day.

Pedagogical inventions like two column proof (Herbst, 2002) or the canonical method for solving equations (Buchbinder, Chazan & Fleming, 2015) support, among other functions, teachers’ management of instructional exchanges by formatting student work in ways that make it more easily assessed with a quick survey by the teacher. With these sorts of pedagogical inventions, each piece of student work does not need its own careful hermeneutic interpretation. But, while such inventions may help teachers manage instructional exchanges, at the same time, these sorts of inventions are often criticized for misrepresenting the discipline (Schoenfeld, 1988) or stunting student learning and removing opportunities for student creativity (Rittle-Johnson & Star, 2007). Perhaps instruction could be changed if a large amount of the responsibility for managing the instructional exchange were off-loaded onto technology, freeing up teachers’ attention for student solutions that are out of the ordinary in different ways. We explore this possibility by first sharing vignettes of current classroom practice and an imagined alternative. We then articulate important characteristics of systems using computer checking of student responses that are intended to support teachers’ formative assessment in classrooms.

Illustrating and articulating the potential of technological supports for assessment in the service of instruction

Technological developments have made it feasible to assess students’ classroom mathematics work automatically and offer important affordances for assessment in the service of instruction when students are working in rich digital environments. In this section, we present two vignettes to outline the potential of such technological support.

Vignette 1: current practice

To begin, imagine a high school mathematics teacher halfway through the academic year teaching calculus to a group of students and focused on tangents to functions. In particular, the teacher wants to learn about students’ concept image of tangents to functions and to assess whether students think that the same line can be tangent to a function at two different points. The teacher also wants to understand how proficient the group is in providing several examples that meet such a given mathematical condition, and which characteristics of the functions change over the different solutions. The teacher could simply ask a related question in class and use student responses to gauge what students know or do not know. But, such a procedure limits the teacher’s
information to students who volunteer or who are called on to present their ideas. Alternatively, the teacher could pose a task for students to do independently, either for homework, or in the beginning of a class period as a warm-up. For example, the teacher might give the following task:

**Claim:** There are functions that have one line tangent to their graph in two different points.

If this claim is true, sketch 3 examples that are different from one another and that support the claim. If this claim is false, explain why you think it is not possible to produce an example.

Each student solution for this task will give a teacher information about each student’s understanding. In our current teaching arrangements, however, the work involved for the teacher in analyzing each solution for the processes that are involved in each student’s understanding is quite taxing (Ball, 2001; Schoenfeld, 2011). If the task is given as a homework problem, the teacher must collect all of the student work, examine each student paper, and devise a mechanism for recording what they have learned. If the task is given as a classroom warm-up, there is less time to examine the student work and a teacher’s assessment is often limited to what they can see as they circulate among students; it is unlikely that a teacher can scan the work of the whole class.

In the next vignette, we present an alternative approach, to illustrate how technology might support assessment in the service of instruction and create an environment that is more responsive to student understandings.

**Vignette 2: technologically-supported assessment in the service of instruction**

Imagine instead a class in which every student has access to a tablet, smartphone, or computer that can run an interactive diagram, one state of which is given in Figure 1, and that the teacher poses the following problem as a warm-up:

**Claim:** There are functions that have one line tangent to their graph in two different points.

If this claim is true, in each of the 3 diagrams shown in Figure 1, provide examples that are different from one another and that support the claim. Sketch the graph of the function with a tangent line and highlight the points of tangency to the graph. If this claim is false, explain why you think it is not possible to produce an example.

This task is designed within the Seeing the Entire Picture (STEP) formative assessment platform [1] so that with the interactive diagram, students can enter an expression for a function or sketch a graph with a pen tool. They can then create a line and drag it to be tangent to the graph at two points. As students work, they can save as many states of the interactive diagram as they wish [2]. When they have a state of the diagram that they think fits the conditions, they can move to the next example, and finally submit that state as a solution to the task. If students wish, they can submit multiple solutions.

Now, imagine that this system collects all of the submitted examples created by all of the students in real time and is able to show the teacher the entire picture of all of the submissions as examples of functions with a line that is tangent to the function in two places in a “carpet” of solutions (Figure 2).

Imagine that apart from the ability to provide a report with assessment information about every individual student, the system can also filter the entire collection of submitted examples to present two kinds of submissions that are not actually supporting examples for the claim: functions that do not demonstrate two points of tangency (e.g., any example with a parabola as the graph of the function), or submitted examples that do not represent functions (see Figure 3).

The teacher in our vignette was not surprised by these incorrect examples, but was gratified to see that a substantial majority of the students submitted examples that fit the criteria in the claim. Looking at the correct submissions, two types of examples stood out to the teacher: two examples representing functions with an asymptote (Figure 4) and three submissions in which the function seemed to be a straight line (Figure 5, shown on p. 14).

After explaining why the incorrect responses had examples that did not illustrate the claim, the teacher decided that the unexpected correct examples with linear functions should be shared with the class. In a whole group discussion format, the teacher asked the students who had given these responses to explain their thinking. The teacher was curious to see whether all students would agree that there are lines with two points of tangency to such a function. The teacher’s hope was that such a discussion would create opportunities to consider whether such “extreme” examples with more than two points of tangency should count or not and why.

**Contrasting approaches to formative assessment**

We contrast the sorts of formative assessment in the classroom that are described in Vignette 2 with current emphases on finding patterns in large datasets of summative assessments that are now often identified with the term “big data.” Current big data initiatives focus on holding educators accountable for student results (Johnson et al., 2014;
Figure 2. Student sample: submitted examples from the STEP platform.

Figure 3. Incorrect examples for the claim from STEP platform data.

Figure 4. Interesting examples with asymptotes.
McCallum et al., 2003). The analysis is about relationships between data on student work (mostly whether it is right or wrong), data about the nature of the assigned tasks, and metadata about students and the context of instruction. The nature of the data on student work makes it difficult to understand how such systems could provide teachers with information that could be used to improve instruction (Dyckhoff et al., 2012). By contrast, the story in Vignette 2 suggests a stronger set of metadata more directly related to student work—metadata based on computer coding of relationships between submitted student work and characteristics of the task. With such metadata, the sort of system implied in the vignette allows teachers to monitor the types of work their students submit, supports teachers in identifying patterns in student errors and correct work, supports teachers in identifying efficient student heuristics for searching a problem space in a task, and provides resources to teachers for communicating norms of what constitutes an interesting or complete response to a task (Figures 3, 4, and 5). As the mediator of the submissions and filters, one might even think of the teacher as a filter of filters through which the class’s work is distilled and understanding is furthered [3].

**Some design considerations for such a technological system**

We have been actively engaged in designing the STEP platform to support the sort of real-time, classroom formative assessment of inquiry learning described in Vignette 2. To describe the system and our rationale for its design, we summarize three phases in the formative assessment process that we see as crucial. Along the way, we describe seven key characteristics of the particular system we have designed.

**Student creation and submission of examples for provided tasks**

Central to technological support for teacher formative assessment is computer assignment of metadata to student submissions that allow both for automatic coding of student responses and the creation of categories among student submissions. Such a technological environment is, therefore, predicated on having students do their work with a digital tool that allows students to create both examples and non-examples of mathematical objects discussed in assigned tasks. We imagine that students use what Naftaliev and Yerushalmy (2013) call an interactive diagram to explore an example space and submit examples that respond to a prompt given in a task. Thus, the formative assessment environment is based upon tasks and digital tools that are in a particular relationship. Not all digital tools and mathematical tasks will work.

First, computer scoring is predicated on the creation of exploratory tasks that have unambiguous criteria for determining correct student submissions. One important genre of such tasks is the sort illustrated in the work of Leung and Lee (2013). In such tasks, which might be called construction tasks, a mathematical object (such as a geometrical object or particular functions that share a particular characteristic), is described and students must produce an example of such an object (see Figure 9, shown on p. 16).

Another kind of task that also requires the construction of examples (Nagari-Haddif & Yerushalmy, 2015) asks students to argue against the truth of a mathematical claim by constructing a counterexample (proving that the claim is false) or constructing an example as evidence for an existence claim (as students are asked to do in the Tangent task in Figure 1). So, while Leung and Lee (2013) are correct to point out that not all tasks can be computer scored, it seems important to continue to explore a variety of task genres that could be used in such environments.

Second, the digital tools that students use to create examples must come from applications that support exploration of mathematical objects [4]. The environment presents an interactive diagram, usually appearing with a non-example that has to be changed within given constrains to reflect examples or non-examples.

Third, to indicate explicitly to students that there is a distinction between their exploration and the examination of their work carried out by the system and to respect students’ intellectual privacy, in the STEP system, students must submit an example in response to a question. By submitting an example, students indicate that they wish to have their submission analyzed as a solution to the task, triggering the use of the unambiguous criteria associated with the task. For this reason, the system contains a mechanism for students to save, revise, and submit their work when they believe that they have completed a solution that the computer can analyze. This process also creates the possibility of having students submit multiple solutions for a given task and to ask students to rank their submissions along different dimensions (e.g., the most surprising example I found), thus enhancing students’ meta-cognitive skills by reflecting on their work. Additionally, requesting additional examples

---

Figure 5. Examples of a straight line as the function.
helps teachers to gain greater insight into students’ thinking. When students are asked to submit examples that differ from one another, it often happens that across all the students in the class, there is greater variety among the second or third examples than among the first examples. Other systems might make other choices that could make it possible to follow student solutions during an exploratory phase (e.g., where did the student drag a point before submitting? What kind of numerical values did the student check as examples to satisfy an equation? Which values did a student try as values in a table of values of a function?).

**Interpretation and storing of submissions**

In the image of the kind of formative assessment tool we are proposing, student submissions are what the computer interprets for the teacher. There are two types of metadata that support communication of the mathematical characteristics of student submissions. The tool in which the interactive diagram was created assigns the first type of metadata. Beyond task solution criteria, this metadata includes sets of mathematically defined descriptions of the submitted work. The second type of metadata is assigned by the platform, to indicate how the submitted examples fit into categories of solutions.

For example in the STEP platform, students work in GeoGebra. As implied in the vignette, for each submitted solution, Geogebra provides the platform with metadata that includes the slope of the tangent, the number of local extremum points of the function, and asymptotes (as an indication of a discontinuous graph), etc. But, GeoGebra can also provide much more information, such as the length of segments, or the position of points and segments relative to each other in a geometry task.

However, just interpreting the solutions is not enough; a system to support formative assessment must store submitted student work in a database for inspection, creating an opportunity to off-load part of the responsibility for the management of instruction onto the technology [5]. Teachers can view student submissions by class or for all their classes, as presented in the vignette (Figure 2) [6].

**Presenting sets of student submissions to their teachers**

As Black and Wiliam (2014) suggest, the feedback that teachers give students on their written work is a central form of dialogue that can promote formative assessment interactions. Such interaction requires that teachers receive more information about their students’ work than simply the possibility of viewing it, or simply knowing whether student work is right or wrong. For a formative assessment platform, supporting teachers involves adding additional metadata to student submissions and designing ways to represent large amounts of data to teachers. For example, we want to bring together the interactive diagram used by students with the database function to create new possibilities for helping teachers understand the nature of the space of examples submitted by many students. We imagine that by using the mathematical properties of student submissions provided by the interactive diagram, additional metadata can be attached to each student submission to indicate its membership or non-membership in a group of submissions. We call these sets of mathematical properties that define membership in a group of submissions filters, with the image that a filter only lets pass the submissions that are members of the specific group, while blocking all other submissions. An important default filter involves whether the submission is correct or incorrect. This filter is an automatic scoring function. But filters are not limited to automatic scoring. In addition, we would like all of this data to be available to teachers in real time; as students submit their work, we would want the representations of the sets of submissions updated in real time for the teacher’s classroom use, thus saving organization time for the teacher by offloading the task of collecting and arranging the assignments for subsequent classroom use.

The following example illustrates how such support for teachers’ examination of student work has been operationalized in the STEP platform. In the basic task illustrated in Figure 6, students were asked to interact with and change the diagram so that the ratio of the areas of triangles ABC and NMC is 1:4. Students had access to the measures of the corresponding sides of the triangles (sides AB and MN).

The four sample student submissions shown in Figure 7 (overleaf) indicate the types of responses we received when piloting this task. For example, solutions I and IV are correct, and II and III are incorrect. But many other filters are possible. In solution III, the submission confuses the ratio of lengths with that of areas in similar triangles, possibly an indication that the student who submitted this example believes the ratio between the sides of similar triangles to be the same as the ratio between their areas, whereas in reality, the ratio of the areas equals the ratio of the sides squared (we call the related filter “a familiar mistake”).

There are other characteristics that hold for both right and wrong solutions, and can be a basis for a filter. For example, solutions I and II use whole numbers, suggesting that students might have used dragging to adjust the image on the screen to have side lengths that are whole numbers, a key option afforded by interactive diagrams (Naftaliev & Yerushalmi, 2013). We call the related filter “easy calculations”. Similarly, solutions I and III involve triangles in which at least one side is horizontal or vertical, an orientation that produces a triangle with what some researchers call a prototypical shape (Rosch & Mervis, 1975). We call the related filter “horizontal/vertical orientation”.

![Figure 6. Initial state of the interactive diagram, to be changed by dragging in the STEP environment using GeoGebra.](image-url)
We now think of tasks in the STEP platform as also involving the creation of filters to help categorize student responses. Such filters can be developed over time and experience with a task. The system cannot provide an exhaustive list of filters, but teachers can also simply select individual submissions of interest. When considering the viability of this platform in everyday teaching, a long-term challenge is to understand what kinds of filters are most helpful to teachers and what sorts of filters might be useful across tasks to allow teachers to track how their students’ responses are changing.

Similarly, we have wrestled with helping teachers understand student work in real time, by providing new ways of representing the range of student submissions to a problem, in addition to the complete set of submissions and the subsets created by the filters. For example, the platform dynamically constructs Venn diagrams representing how the submitted examples populate the filters that exist for a task, as well as the relations between problem filters (Figure 8). Of course, this sort of representation can only be explored for a limited number of filters at a time.

Another representation of the range and type of student submissions is illustrated below. In the construction task shown in Figure 9, adapted from Chazan (1989), students are able to drag point A while points B and C are fixed. They are asked to create a right triangle, and to submit 10 different examples.

In Figure 10, we show two aspects of reports for this task: a representation of all of the locations for point A that students have submitted (on the left) and a representation of the locus of all possible locations of A that will create a right triangle (on the right). Examination of the two, side by side, suggests different incorrect patterns (e.g., isosceles triangles). These reports are presented within GeoGebra and thus point A can still be dragged in real time to explore them. Locations are updated in real time and can be used by teachers as their students work.

To summarize, in order to have students be able to create and submit examples for given tasks, to have the platform be able to interpret and store their submissions, and to present information on sets of student submissions, seven key characteristics were identified and designed. These characteristics are portrayed on the line at the center of Figure 11.

Conclusion
In this article, we have attempted to illustrate how technological advances might be harnessed to support the work of teachers by providing them with immediate access to views of the work of large groups of students working on open response tasks, categorizing them to fit relevant characteristics for the teacher to attend to, and also freeing up teachers’ attention for student solutions that are out of the ordinary in different ways. Such supports might allow for better classroom representation of the discipline of mathematics, more effective student learning, and greater opportunities for student creativity. For us, important next steps include gaining substantial experience about the use of such a system by teachers in a range of contexts. We seek to understand the instrumental genesis of such a tool for teachers, the impact of the use of such a tool on classroom interactions, and ultimately its impact on student learning of mathematics. We hope to see the mathematics education community build on these ideas to develop technological tools that support teachers to quickly survey and understand student solution strategies for open-ended tasks in a digital environment, and to use these understandings in the service of instruction.

Notes
[1] Seeing the Entire Picture (STEP) is a formative assessment platform developed at the University of Haifa’s Center for Research and Development of Alternatives in Education. Work with this platform has shaped our ideas about the use of computer checking of student work for the purposes of formative assessment. For more detail about this platform, see www.viustep.com.
[2] This vignette is based on task design and classroom experiments with
the STEP platform that are part of Galit Nagari-Hadiff’s PhD dissertation (in progress).


[4] In the case of the STEP platform, this environment is GeoGebra (Hohenwarter & Preiner, 2007).

[5] The current version of the STEP platform is consistent with the standard and format used by Moodle for organizing and storing responses.

[6] For a similar application that enables only the viewing of student and of whole-class work with the TI-NSPIRE, see Clark-Wilson (2010) and Arzarello and Robutti (2010).

References


