# Talk Aloud Problem Solving and Frequency Building to a Performance Criterion Improves Science Reasoning

*Ginny A. Dembek*
*Richard M. Kubina Jr.*

ABSTRACT: The present study examined the effects of a combined intervention: Talk Aloud Problem Solving (TAPS) and Frequency Building to a Performance Criterion (FBPC). The experimenter introduced TAPS/FBPC to five students diagnosed with a disability and receiving specialized reading instruction. The intervention presented TAPS formatted lessons and FBPC strengthened the student's verbal repertoire making the problem-solving process a durable behavior. A multiple baseline design showed improvements in problem-solving performance when compared to baseline. All students became more accurate in the problem-solving task, as shown in immediate changes upon the implementation of the intervention and sustained growth over time. Maintenance in learning was also demonstrated. Implications for practice and future research are discussed.

KEY WORDS: Talk aloud problem solving, Frequency building, Behavioral fluency, Problem solving, Science education

## THE PROBLEM

Science has begun to receive the attention it deserves through mandates and other important legislation. Despite the increased attention on scientific literacy, students with disabilities by 8th grade have greater difficulties compared to their peers with 66% performing below basic according to the most recent National Assessment of Educational Progress science results (NAEP, 2015). Although scientific literacy remains important, reading skills prior to comprehension

Address correspondence to Ginny A. Dembek, Ph.D., Brooklyn College, CUNY, Department of Childhood, Bilingual, and Special Education, 2900 Bedford Ave., 2205 James Hall, Brooklyn, NY 11210. E-mail: gdembek@brooklyn.cuny.edu.



JEBPS 16(2).indb 147 8/21/2018 7:55:45 PM

already require more intense instruction. Students with learning disabilities demonstrate more severe forms of reading problems in comparison to poor readers not identified as having a learning disability (Fuchs, Fuchs, Mathes, & Lipsey, 2000).

As the reading materials become more difficult, especially in content areas using expository text, growing numbers of students require more intense and effective instruction to fully benefit from both in-class and out-of-class activities. Expository text introduces many challenges and complexities increasing difficulty for content area learning. Students with disabilities, for example, receiving intense reading instruction as per their Individualized Education Plan (IEP), would still be expected to perform with some minor accommodations to solve problems in content areas such as science. Expository text brings additional challenges with advanced vocabulary, varied and complex text structure, conceptual density, and a requirement for additional prior knowledge (Saenz & Fuchs, 2002). With more complex reading materials, additional instruction is required.

Based on what traditional instruction includes, students are not receiving the instruction that is compatible with their needs (Brigham, Scruggs, & Mastropieri, 2011) or necessary to promote the understanding and comprehension of complex expository text (Mason & Hedin, 2011). Science requires students to continue to build knowledge representing a current understanding of natural systems and the process whereby that body of knowledge becomes established and continues the process of extension, refinement, and revision (Duschl, Schweingruber, & Shouse, 2007). The process of science, like reading, requires a solid foundation of skills in order to become scientifically literate.

## PREVIOUS RESEARCH

The difficulty of comprehending more complex types of text can be lessened by teaching students to problem solve. One intervention developed by Whimbey and Lochhead (1999) that has shown promise with older students (i.e., high school and college) is Think Aloud Pair Problem Solving, also referred to as TAPPS (Glass, 1992; Holzer & Anduret, 2000; Jeon, Huffman, & Noh, 2005; Johnson & Chung, 1999; Kani & Shahrill, 2015; Pate & Miller, 2011; Pate & Young, 2014; Pate, Wardlow, & Johnson, 2004; Pestel, 1993; Tingle & Good, 1990). The structured TAPPS intervention includes modeling and prompting of think aloud statements. Implementation can take place across skills and content areas while working in pairs. Research with talk alouds, also called think alouds, during reading has demonstrated the importance of including explicit modeling and prompting with instruction (Baumann, Seifert-Kessell, & Jones, 1992; Bereiter & Bird, 1985; Witcoski, 2012). An adaptation of the broader intervention of TAPPS, Talk Aloud Problem Solving or TAPS has features designed specifically for an individual student and can be implemented with pairs of students as they master the process of TAPS. Most importantly, the student benefits by becoming his or her own listener and more independent (Robbins, 2011).

TAPS serves as a good tool for building accuracy in problem solving. However, effective practice is required for students to become proficient in problem solving. One way to replicate the skill of thinking aloud while solving problems is adding a frequency building component. Frequency building requires a student to repeat the process of thinking aloud in a timed practice trial, receiving performance feedback, and practicing again (Kubina & Yurich, 2012). The goal of frequency building is to practice until proficient at a skill, achieving behavioral fluency. Adding a criterion to the performance provides students with an end goal and helps establish literature for behavioral fluency with different academic behaviors. Frequency building to a performance criterion (FBPC) follows previous research indicating that once a behavior reaches a frequency standard (a threshold demonstrating behavioral fluency), behavior will maintain across time after instruction has ended (Binder, 1996; Kubina, 2005; Kubina, Amato, Schwilk, & Therrien, 2008; Kubina & Yurich, 2012).

A recent study extended the TAPS/FBPC literature by implementing the problem-solving strategy using science text in fourth grade. Dembek and Kubina (in press) used TAPS combined with a frequency building component with a student in fourth grade at-risk for learning disabilities who demonstrated difficulties in reading comprehension and problem solving. The problem-solving task of placing sentences from a science textbook (i.e., four sentences out of order) in a logical order was selected as the dependent variable. The task chosen required students to organize a passage while demonstrating their awareness of text structure. The independent variable was the TAPS/FBPC. The intervention package included explicit instruction lessons (model, guided practice, and check) and timed practice with feedback. The experimenter found that TAPS/FBPC demonstrated a positive effect on the student's ability to place sentences in a logical order.

## THE SOLUTION

The positive effect seen in the TAPS/FBPC study reflected the student's ability to problem solve and determine a logical order of content. The student used text structure, concepts, vocabulary, and prior knowledge to help them, all areas making science text a more difficult text to read (Saenz & Fuchs, 2002). Dembek and Kubina (in press) began to address a previous concern suggesting students with learning disabilities may not have as much awareness of passage organization (i.e., text structure) and struggled reorganizing disorganized passages when compared to students without learning disabilities (Wong & Wilson, 1984). Strategies helping students navigate text structure (e.g., TAPS) and apply appropriate structure-specific strategies may benefit learners across varied prose (Bakken, Mastropieri, & Scruggs, 1997) and science (Dembek & Kubina, in press).

One way to engage the language of problem solving in science is through TAPS. TAPS has been successfully implemented in high school and college settings as well as with one student in fourth grade. Students who practiced the language of problem solving to explain their process like older advanced students (i.e., students in advanced placement science courses in high school), stated more details in explanations. However, the fourth-grade student in Dembek and Kubina (in press) was at-risk for failure in literacy and not identified with a learning disability.

Students with learning disabilities, included in the general education curriculum, need to learn to vocalize their process of problem solving and self-advocate for additional instruction, clarification of text, or additional accommodations. The TAPS intervention fosters a skill that students with learning disabilities can benefit from in science. To explore TAPS with students with learning disabilities in reading, the following questions were asked: (1) To what extent does the combined intervention of TAPS and FBPC impact students with learning disabilities ability to reason or problem solve content from a science text? (2) Do students maintain the ability to solve problems with science text over time following TAPS/FBPC intervention? (3) Does the problem-solving skill targeting science content generalize to other content areas, specifically social studies? (4) Do teachers and students view the intervention of TAPS/FBPC intervention as a socially valid and beneficial classroom tool?

JEBPS 16(2).indb 150 8/21/2018 7:55:45 PM

## EVIDENCE OF EFFECTIVENESS

### Participants

The experimenter outlined specific procedures to identify participants who benefited from inclusion in the TAPS/FBPC study. First, teachers nominated students in their classroom that struggled with test-taking and reading comprehension skills. Following nomination and parental consent, each student was assessed for their current oral reading fluency and retell ability. The experimenter identified students who struggled with comprehension not due primarily to decoding problems. Students with a low reading frequency or rate, (less than the 10th percentile for winter on grade level) specifically below 61 correct words per minute (CWPM), may not benefit from the intervention because of the need for more intense instruction in decoding. The oral fluency norms were obtained from Hasbrouck and Tindal (2006).

Along with decoding fluency, students retold what they remembered after reading a one-minute passage. The students needed to retell five or less independent clauses (i.e., can stand alone as a sentence and contains a subject and a verb) per passage in one minute based on half of a fluent retell performance (Culler, 2010). An example of a sentence that would count as two independent clauses, or two retells is the following. The car drove down the street/and she waved goodbye. The reading fluency and retell scores needed to occur in two out of three passages in order for the student to qualify for the study. Table 1 provides specific participant information.

### Setting

The study was conducted in a suburban/rural public school district located in the northeastern United States. The delivery of the intervention took place at the convenience of the teacher and students in two separate elementary schools. Typically, instruction occurred in the learning support classroom or in the hallway directly outside of the classroom. Depending on student schedules, the student received instruction during the learning support classroom or general education classroom content for the study. The majority of days, the intervention occurred during learning support instruction. The total enrollment of the school district for 2012–2013 school year was approximately 6,700 students. The school district provided special

**Table 1. Student Characteristics**

| *Variable* | *Level* | *Students* | | | | |
|---|---|---|---|---|---|---|
| | | *Abbie* | *Jada* | *Saddie* | *Sasha* | *Theodore* |
| Demographics | Gender and race | Female Caucasian | Female Caucasian | Female Caucasian | Female African American | Male Caucasian |
| | Exceptionality | SLD in reading and writing | OHI due to ADD | SLD in reading | SLD in reading OHI | OHI due to fetal alcohol syndrome |
| Grade Level Reading Fluency | CWPM/IWPM reading one | 87 CWPM/5 IWPM | 101 CWPM/4 IWPM | 65 CWPM/4 IWPM | 89 CWPM/3 IWPM | 101 CWPM/3 IWPM |
| | CWPM/IWPM reading two | 71 CWPM/6 IWPM | 79 CWPM/4 IWPM | 61 CWPM/6 IWPM | 79 CWPM/5 IWPM | 76 CWPM/3 IWPM |
| | CWPM/IWPM reading three | 52 CWPM/6 IWPM | 59 CWPM/5 IWPM | 57 CWPM/5 IWPM | 88 CWPM/4 IWPM | 94 CWPM/1 IWPM |
| Retells | Correct and Incorrect reading one | 6 correct/0 incorrect | 1 correct/0 incorrect | 4 correct/0 incorrect | 2 1/2 correct/0 incorrect | 2 correct/0 incorrect |
| | Reading Two Correct and incorrect | 2 correct/1 incorrect | 2 1/2 correct/0 incorrect | 4 correct/0 incorrect | 2 1/2 correct/0 incorrect | 3 correct/1 incorrect |
| | Reading Three Correct and incorrect | 2 correct/1 incorrect | 1 correct/0 incorrect | 3 correct/0 incorrect | 2 correct/0 incorrect | 3 correct/1/2 incorrect |

*Note*: ADD = Attention Deficit Disorder; CWPM = Correct Words per Minute; IWPM = Incorrect Words per Minute; OHI = Other health Impairment; SLD = specific learning disability.
MD_12

education to approximately 785 students with 20% of all students qualifying for free/reduced lunch. The school district demographic information shows 84% of students identifying as Caucasian, 3% of each the following identifying as Hispanic, mixed race, and African American, and 7% identified as Asian.

### Materials

Materials consisted of paragraphs from science textbooks used in fourth-grade classrooms. Some materials had already been developed from textbooks for fourth-grade students (Banks et al., 2001; Cooney et al., 2006; Dawson-Boyd et al., 2006; Hacket et al., 2008; Heil et al., 1994; Sciencesaurus, 2002) and used in a previous study (Dembek & Kubina, in press). Papers were provided to the students with the paragraphs as well as timers and a pencil or pen. A voice recorder was used to record the student responses. Each page of the dependent variable contained three to four sets of problems (Figure 1 displays a sample of one problem). Each problem contained four sentences that could be linked consecutively. The student had the possibility of 16 correct responses on each page (when containing four problems), with another page available if needed.

### Dependent Variable

Students answered a basic problem-solving task related to reading comprehension (Whimbey, 1995). The task, presented as four sentences, required the student to read each sentence and write the correct order in a two-minute timed session. The order was written using a number from 1 to 4 next to the sentence which indicated the students selected answer. The dependent measure was scored using the correct sequence of sentence or CSS and incorrect sequence of sentence or ISS. Students completed the problems on each sheet and received another sheet if all problems were completed. The selected response was scored as correct or incorrect based on the order provided in the textbook. A sample of a problem included in the dependent variable appears in Figure 1.

### Experimental Design

The experimenter used a multiple baseline experimental design across participants (Gast, 2010). A multiple baseline design across participants remains useful for skills that cannot be unlearned, such as many academic skills (Cooper, Heron, & Heward, 2007). The multiple baseline design demonstrates strong internal validity if a functional relation is demonstrated between the independent and dependent variables in a study (Kennedy, 2005). All data are displayed on segments from the Standard Celeration Chart (SCC).

### Accuracy of Dependent Variable and Procedural Integrity

The experimenter calculated accuracy to find the extent to which observed values estimated the events that took place in an experiment (Johnston & Pennypacker, 2009). She scored the dependent variable, writing the correct sequence of sentences. As an additional measure to score more accurately, the experimenter scored all of the problems twice. An independent scorer re-scored 33% of the dependent variable to check for accuracy. The scorer followed an answer key to check the answers (i.e., training was simply teaching the scorer how to use the answer key with problem codes). While rescoring the dependent variable, the independent scorer wrote what the scoring should be and what it was.

The observers were a fourth-year special education major and honors student and a school psychology graduate student. The independent observer went through the answer key and checked 33% randomly selected dependent variable probes demonstrating 99% accuracy in scoring. Along with the order or answers, another scorer checked a new scoring system with 99% accuracy. The second scorer checked the correct transfer of the student's written problem order as well as the score with a scoring guide.

One independent scorer also listened to 33% of the lessons and determined the procedural integrity. The scorer used the scripts the experimenter implemented in the study and followed along with the audio version. When the scorer heard specific content included in each paragraph, she checked off the area in front of the part of the lesson.

An additional measure, the scoring of the independent variable measure (i.e., timed talk aloud prompt), improved the procedural integrity. The calibration, or evaluation of data from the measurement procedure, helped to adjust and improve the procedure (Johnston & Pennypacker, 2009). Training was required for the scorer in order to reach 100% agreement on at least three examples. To calculate reliability, each individually scored probe was calculated, and then all of the probes were combined to find the average. The experimenter randomly selected 33% of the talk aloud transcriptions and reliability came to 98%.

## Social Validity

The experimenter provided teachers with copies of questions for each student, as well as a sheet with teacher questions. Students were asked by their classroom teacher after the completion of the study to avoid pressure to answer the questions positively in front of the experimenter. The questions included: First, do you feel the instruction helped you in school? How? Second, did you like the time during school you were pulled out of your day for instruction in talking aloud? Third, do you think this instruction could help other students your age? Why? Fourth, what do you think could be added that would help you more? And the last question, would you want to work on this skill again?

The teacher questions included: First, do you feel like this intervention is beneficial for students with disabilities? In what area? Second, have you noticed a difference in any other content? Third, should we have chosen this behavior as a goal for the intervention? And the last question asked to teachers was, if the probes and intervention were already made, could you see yourself adding this to your classroom instruction if time was made available?

## Experimental Procedures

### Baseline

Students began baseline after parental consent and initial testing. The experimenter measured student behavior in baseline for at least five data points (What Works Clearinghouse, 2011) using the dependent variable (i.e., same directions as during the intervention). Similar to a previous study (Dembek & Kubina, in press), the student who began the intervention first had the most substantial deterioration, or evidence of no growth, of accuracy over time. Once the first student reached the criterion for TAPS, the next student with a stable set of learning data began the intervention. The second through last student entered the intervention when the student before them moved onto the second stage of the intervention.

### Talk Aloud Problem Solving (TAPS)

The first components of the TAPS/FBPC were 10–15-minute lessons on talking aloud the logical order of sentence. Following the scripted lesson, a two-minute timed talk aloud prompt and then the dependent variable were completed. The two-minute timed talk aloud prompt consisted of one, four-sentence prompt similar to the dependent variable. To determine student success in the lessons, the two-minute timed talk aloud prompt was recorded and categorized.

The three categories of talk alouds included rereading, talk alouds about order, and talk alouds about cue. Rereading required point-to-point correspondence to the text. Students scored one point for each of the sentences (e.g., the fox ran down the road) or 1/2 point for part of the sentence (e.g., the fox). Talk alouds about order were any independent clause about order without any explanation (e.g., this sentence goes first). In other words, the talk alouds about order only stated a position the sentence was arranged in. A talk aloud about cue included any independent clause that used part of the sentence and words to give explanations for decisions (e.g., this sentence explains the other sentences). The talk alouds about cue included anything that lead the student to get to an answer including statements of misunderstandings or unknown information (e.g., I do not know what this word means, I am not sure where this sentence goes so I should keep reading).

When a talk aloud about cue combined with a talk aloud about order, it counted as two talk alouds about cues (e.g., student says, "The sentence I just read will go last because it explains the other sentences"). A talk aloud counted as incorrect if it was unrelated to the content in the sentences or unrelated to problem solving. During the TAPS lesson stage, no feedback followed the two-minute timed talk aloud prompt. The TAPS lesson stage ended when students reached half of the exit criterion of the study (i.e., eight talk alouds about cues). The goal of TAPS was to instruct students on talking aloud about their decision-making and increase explanations.

### Frequency Building to a Performance Criterion (FBPC)

The second stage of the intervention began when students could profit from practicing the skills learned correctly during the initial TAPS instruction. To enter the second part of the intervention the student talked aloud about cues eight times in two minutes (i.e., half of the exit criterion of the study). The goal of the FBPC was to practice problem solving with feedback. FBPC involved repeatedly practicing a timed behavior followed by immediate performance feedback after the practice trial (Kubina & Yurich, 2012). The student continued practicing over time until he or she reached a performance criterion or quantitative marker predetermined by the experimenter as developed in a prior study.

Students began FBPC with the two-minute timed talk aloud prompt. Following the two-minute talk aloud, one minute of performance feedback was provided and the student repeated the task again. Feedback began with praise recognizing something the student added to their talk alouds previously not observed or something correct (e.g., recognizing that they need to reread a sentence before making a decision). Feedback on incorrects, missing information, or partial answers followed for the remainder of the minute by the experimenter through modeling or asking the student to produce the correct response. The dependent variable was administered following the second one minute of feedback. The students exited FBPC once they reached 16 talk alouds about cues. The frequency criterion became established by previously collected talk aloud data from two "expert" learners in science (i.e., two high school students enrolled in Advanced Placement or AP courses in science) and implemented in a previous study (Dembek & Kubina, in press).

### Maintenance

The experimenter administered the dependent variable for two minutes once every two weeks for one month following the first week check. The extent of feedback included only praise for completing the dependent variable. After completion, the student returned to the previous classroom setting.

## RESULTS

Results from the experiment are presented in graphical format. The authors displayed data on five tiers taken from SCC segments (Pennypacker, Gutierrez, & Lindsley, 2003; White, 1986). The SCC-derived figure allowed the experimenter to illustrate learning that occurred in real time accounting for all days the student was present and absent due to illness, weekends, or holidays using successive calendar days on the horizontal axis. The count per minute frequencies on the vertical access with a ratio scale display each distance on the vertical access of the chart in equal ratios of change (e.g., distance from 1 to 2 is the same distance as 10 to 20, both show a doubling of change). The SCC helps aide in visual and quantitative analysis and facilitates better decision-making for educators.

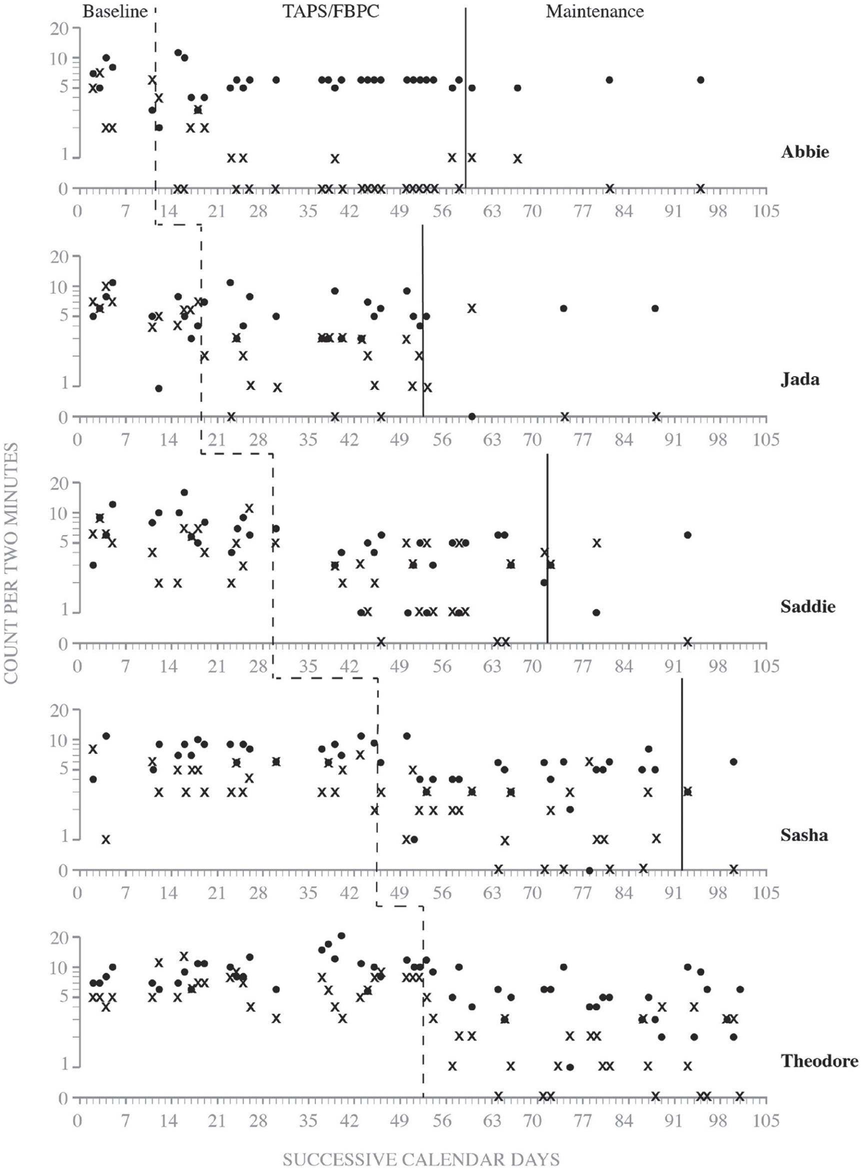


**Figure 1. Sample Problem-Solving Prompt.** One problem was presented to the student for the talk aloud prompt. The student was presented with three to four problems on a page for the dependent variable (as many as needed to fill the two-minute timing).

_____A. They divide each of these groups into even smaller groups.

_____B. Scientists divide kingdoms into smaller groups.

_____C. They continue sorting into smaller and smaller groups.

_____D. Each time they sort, they use the organism's features to decide whether or not the organism belongs to the group.

Using celeration, the significance of behavior change is demonstrated by daily performance frequencies changing over time (Johnston & Pennypacker, 2009, Pennypacker et al., 2003). For example, a x1.0 celeration means that the student's behavior has remained consistent or the trend is flat and not changing. If the goal is improvement, or acceleration, something needs to be altered, such as the implementation of an intervention. The experimenter calculated a trend line by using the quarter-intersect technique (Pennypacker, Gutierrez, & Lindsley, 2003).

Another SCC measure used to analyze behavior change is the frequency multiplier (i.e., also known as frequency jumps), or the amount of change between two frequencies. Single case design calls for examining the immediacy of change when an intervention is applied (Cooper, Heron, & Heward, 2007; Kratochwill et al., 2013). A frequency jump value (i.e., jump up, jump down, no jump) is found by measuring the distance from the first frequency data point to the second frequency data point (Pennypacker, Guitierrez, & Lindsley, 2003). As an example, going from 1 to 2 would represent a x2.0 jump up, or a doubling of performance. Figure 2 displays all student data for baseline, intervention (TAPS/FBPC), and maintenance phases with a counting time of two minutes.

**FIGURE 2: STANDARD CELERATION CHART SEGMENTS SHOWING BASELINE, INTERVENTION (TAPS/FBPC) AND MAINTENANCE PHASES FOR ALL STUDENTS.**

IMPLEMENTATION GUIDELINES

1. Scripted instruction in Talk Aloud Problem Solving (TAPS).
   a. Relevance: The teacher explains to the student that they will be learning to talk aloud while they solve problems. The teacher emphasizes how important it is that problem solvers are clear and provide a lot of detail. The students are asked to provide two reasons why it is important to be able to explain how to solve problems. The teacher can prompt them with questions like: How may this help you in class? How may this help you take a test? Or How does this help you in solving problems with multiple steps?
   b. Teacher Model: Teacher models how to talk aloud solving a problem (same type of problem students will solve). A sample part of one scripted model:
      _____ A. He planted a seedling in a pot of soil.

_____ B. About 400 years ago, a Dutch scientist named Jan van Helmont wanted to know how plants meet their needs.
_____ C. After five years, the seedling became a small tree.
_____ D. He watered it regularly.
______A. He planted a seedling in a pot of soil. I don't know who he is. This is most likely not the first sentence. I will wait until they introduce who he is. I also don't know why a seedling is going to be planted in soil.
______B. About 400 years ago, a Dutch scientist named Jan van Helmont wanted to know how plants meet their needs. This seems like a good introduction sentence. I found out who he is, Jan van Helmont. I also found out when this took place. Now I know why he planted a seed, because he wanted to find out this information. I will keep this one in mind for the introductory sentence.

c. Guided Practice: After at least two models, the teacher starts to involve the student in the problem-solving process. The student is asked to read the sentences in the problem and the teacher uses a combination of prompts from telling the student how to approach the sentence, asking questions about the sentence, or reminding the student to look for key words in the sentence.
d. Independent Practice: The student is asked to complete a problem without any assistance. The teacher should expect the student to provide 8 details about the problem during the explanation. After providing 8 details for at least 3 separate problems, the student is ready to move on to the next stage.
e. Remediation: If the student is struggling at any stage of the instruction, the teacher needs to go back to the previous level of prompting (Full model, telling the student how to complete the problem, asking questions, or reminding the student).

2. Frequency Building to a Performance Criterion: the student continues this phase of the study until he/she is able to talk aloud about the prompts/details in the sentence at least 16 times. The teacher should make them do this on at least 3 problems.
   a. Two-minute timed talk aloud while solving a problem.
   b. One minute of teacher feedback.
   c. Second two-minute timed talk aloud while solving the same problem.
   d. One minute of teacher feedback.

## Correct and Incorrect Sequence of Sentences

### Abbie

Figure 2 displays data from Abbie's results across baseline, intervention (i.e., TAPS/FBPC), and maintenance phases for sequencing sentences. During baseline, five data points, Abbie showed a consistent worsening for correct sequence of sentences or CSS with a celeration of ÷8.5 [10 days] and incorrect sequence of sentences or ISS increasing with a x2.3 [10 days]. CSS and ISS were moderately variable in baseline.

Abbie began TAPS/FBPC first out of the participants and displayed an overall change in direction for her correct and incorrect celerations. Her performance in CSS accelerated to a consistent x1.2 [47 days] with ISS decelerating at ÷1.5 [47 days]. Her frequency jump for corrects was x1.3 and incorrects demonstrated an immediate jump down of ÷3.0 from baseline to TAPS instruction. Abbie demonstrated a stable performance of both CSS and ISS during the intervention. Abbie's performance continued to maintain over five weeks after TAPS/FBPC completed. Following the intervention phase, her performance was measured for maintenance. Abbie demonstrated levels similar to the intervention phase with incorrects below corrects (i.e., three weeks and five weeks after the intervention demonstrating six correct and zero incorrect sequence of sentences), her last two weeks with 100% accuracy.

### Jada

Figure 2 displays data from the second student who started the intervention, Jada. Her baseline performance indicated a steady decline in performance with CSS decreasing at ÷2.2 [17 days] and ISS decelerating as well at ÷1.5 [17 days]. Her corrects and incorrects were both moderately variable in the beginning of baseline. After a week of data collection, the ISS began stabilizing and the CSS remained moderately variable. Despite the improvement in incorrects, the errors were still at a high level compared to corrects in Jada's baseline. Persistent, high-level incorrects above corrects prompted the experimenter to select Jada to begin the intervention next.

TAPS/FBPC yielded an improving performance in corrects. Jada's corrects immediately improved, but demonstrated a slight decrease during the intervention phase and incorrects increased slowly across time, respectively, of ÷1.1 [34 days] and x1.3 [34 days]. Jada showed a drop in overall level for ISS as well as a slight increase in level for

CSS. Upon entering the TAPS intervention Jada immediately demonstrated an improvement in performance of x1.8 CSS with a significant jump down of ÷3.5 in ISS. During the intervention CSS and ISS performance was moderately variable. Data indicated a slight improvement in accuracy from baseline when looking at immediate changes and the impact of ongoing performance, with maintenance data that indicated Jada improved her ability to determine the correct sequence of sentences. Despite the lack of clear improvement in accuracy, the trend of corrects remained well above the trend in incorrects for the length of the intervention, opposite of Jada's learning picture in baseline. She scored zero correct and six incorrect one week following the intervention, but then returned to her highest performance during intervention at three and five weeks after instruction, scoring four correct and zero incorrect sequence of sentences both days.

### Saddie

Once Jada entered the FBPC stage of TAPS intervention, Saddie began the intervention. Saddie showed a consistent decline in performance with the CSS decelerating at ÷1.5 [28 days] and ISS maintaining at a slightly declining celeration of ÷1.2 [28 days]. Her performance during baseline for CSS and ISS was variable. Saddie began the intervention as soon as Jada entered the frequency building stage of the intervention.

Saddie started TAPS/FBPC and continued for 42 calendar days as shown in figure 1. During the intervention, her CSS decreased slightly at ÷1.3 [42 days] and ISS decelerating at ÷1.2 [42 days]. Saddie's correct sequence of sentences had a jump up of x1.2 from baseline phase once beginning the first day of instruction in TAPS. She also had a jump down of ÷2.4 in ISS upon entering the intervention phase. Although Saddie demonstrated an improvement, CSS and ISS data showed variability. The frequency multipliers showed immediate and positive change in the performance of sequencing sentences. A week after the intervention phase, Saddie scored one CSS and five ISS and three weeks after scored six CSS and zero ISS, consistent with her variability in intervention. The data following the intervention demonstrated the effects from the intervention maintained for up to three weeks following the termination of instruction.

### Sasha

Sasha continued in baseline for 44 days with a consistent CSS celeration of ÷1.1 [44 days] and in ISS steady at ÷1.1 [44 days]. Her data showed slight variability for both ISS and CSS in baseline. Sasha began TAPS when Saddie reached the criteria to enter the frequency building portion of the intervention.

During TAPS/FBPC, Sasha's correct and incorrect celerations reversed. Her CSS slightly accelerated at x1.2 [48 days] and her ISS decelerated at ÷1.5 [48 days]. Sasha had a substantial accuracy improvement during TAPS but overall eliminated most incorrects during the time in TAPS/FBPC. Although there was a clear trend in the CSS and ISS data, the data remained variable. Her performance on the first day of the intervention slowed with a jump down in both CSS (÷1.2) and ISS (÷3.5). One data point collected a week following instruction demonstrated a performance of four CSS and zero ISS in maintenance.

### Theodore

Theodore remained in baseline for the longest period of time, 52 calendar days. During baseline his corrects and incorrects maintained. His CSS accelerated at x1.2 [52 days] and ISS accelerated at x1.1 [52 days]. Although both were accelerating, his incorrect responses remained at a high level during baseline with a very slight accuracy improvement. Baseline ISS and CSS showed moderate variability. Once Sasha reached the criterion for FBPC, Theodore began the intervention.

Theodore began the TAPS with an immediate jump up in his CSS performance of x1.3 in correct responses. His ISS also had a jump down of ÷1.6. During TAPS/FBPC the overall level of the corrects and incorrects dropped with a steady performance. Theodore had a small decrease in CSS decelerating at a ÷1.3 [48 days], along with a decreasing performance in ISS of ÷1.3 [48 days]. Although a change in levels was calculated, both ISS and CSS were variable. Overall corrects increased and incorrects dropped and remained steady resulting in a more accurate performance. The school year ended and no maintenance data could be collected for Theodore.

## Social Validity

Four of the students were asked questions concerning the instruction, time of day they were pulled out of the classroom, outcomes, and future use of the intervention. The fifth participant's answers were lost by the teacher at the end of the school year. Participants viewed the instruction in talking aloud and solving problems as a relevant skill. Three of the four students stated that the intervention made them better at reading with help understanding bigger words, and what they were reading in the problems. Abbie stated that the intervention helped her take her time and really think aloud about the order. Three of the four stated it had helped them with a certain skill (answer questions better, reading new and different things, and realize mistakes). All four students thought other students would benefit from the intervention (having trouble ordering events, trouble reading, or not understanding word meaning).

Classroom special education teachers were also provided with a set of questions. Both teachers felt it was beneficial with one teacher stating that the intervention helped students who struggle with communicating what they are thinking. She felt that sometimes they need to be given a script and taught how to "think" out loud and given permission to express themselves in this way. Teachers thought talking aloud about problem solving was a relevant goal with one teacher stating verbalizing thinking was a critical life skill (e.g., what students need help with, and what students don't understand so teachers can better serve their needs).

## DISCUSSION

The effect of TAPS/FBPC on a problem-solving skill was measured using a multiple baseline design across participants. All students increased in accuracy of performance from baseline. Prior to entering TAPS/FBPC, students demonstrated inaccurate performance with a variable and high level of incorrects for the length of baseline. Three of the five participants became less accurate in baseline, some slightly (Saddie and Jada) and others significantly deteriorating (Abbie). The other student, Theodore had slightly improving accuracy, but the level of his incorrects remained high in comparison to corrects throughout baseline (i.e., 52 calendar days) meaning he had consistent high levels of incorrect answers when solving problem. Additionally, for four out of five students the baseline for accuracy demonstrated that even with general education science instruction, the students did not improve through classroom instruction or maturation.

The TAPS/FBPC had a positive effect on the quality of reasoning about science content and problem solving. The data reflect similar findings by Pestel (1993) showing that students introduced to instruction answered fewer problems completely right, with the other group getting more problems completely right or wrong. Early stage TAPS/FBPC learners appear to have a tendency to try to work through the problem rather than either knowing or not knowing the answer. By the end of the study students were demonstrating a promising understanding of textual features to help aid in organizing paragraphs as well as the ability to verbalize their process.

The experimenter introduced the TAPS/FBPC to all students, with only four reaching the exit criterion (16 talk alouds about cues). Frequency multipliers were used to quantitatively examine the immediate change in student performance once instruction for the intervention began. The measure demonstrated favorable consistency across students with a drop in incorrects across all participants. Upon introduction to the intervention, Saddie, Abbie, and Theodore all had slight jump ups in corrects (ranging from jump ups of x1.2 and 1.3) with Jada demonstrating a more substantial increase of x1.8 upon entering the intervention. Sasha demonstrated a slight decrease in her corrects of ÷1.2 CSS during the first day of instruction. Despite the smaller immediate impacts on correct responses, the impact on incorrect responses became much more substantial. The ÷2.0 frequency multiplier for incorrects indicates the intervention had a strong immediate impact on students' ability to problem solve. Abbie, Jada, and Sasha all demonstrated substantial jump downs of incorrects of ÷3.0 for Abbie and ÷3.5 for both Jada and Sasha. The other students both showed a substantial immediate decrease in ISS of ÷2.4 for Saddie and ÷1.6 for Theodore. When placed in context, the TAPS/FBPC produced substantial decreases in the incorrect answers for students immediately upon entering the intervention.

The study demonstrated improvement consistent overtime through the celeration measures. The celeration for all students prior to the intervention showed the magnitude of need to begin instruction. Abbie, Jada, and Saddie had significant reductions in correct answers with, respectively, ÷8.5 [10 days], ÷2.2 [17 days] and ÷1.5 [28 days] decelerations. While incorrects were also decelerating, Abbie ÷1.2 [10 days], Jada ÷1.5 [17 days], and Saddie ÷1.2 [28 days], the corrects outpaced the incorrects in loss of learning. Sasha and Theodore both demonstrated that without instruction, the long trends of near flat lines, or no change in inaccurate performance, would continue for a large portion of the school year. The significance for Sasha's corrects, x1.1 [44 days] and Theodore, x1.2 [52 days] both fall in the range of unacceptable growth celeration (Kubina & Yurich, 2012).

The present study also showed benefits in TAPS/FBPC over time. All student performances during the intervention stayed consistent or remained better than in baseline. With a multiple baseline design, experimental control is determined when an intervention is applied to a tier and no spillover of effects occurs in subsequent tiers (Kennedy, 2005). The intervention must also show significant effects for changes in trend, immediacy of impact, and consistency in data pattern with subsequent phases (Kratochwill et al., 2013). The present study shows a positive immediate impact for all students when introduced to the TAPS/FBPC. Furthermore, the celeration for correct and incorrect sequences of science content improved for two students substantially and three students slightly compared to baseline performance. TAPS/FBPC demonstrated a positive change for all five students with the changes from baseline produced by the TAPS/FBPC indicate an experimental effect.

The experimenter collected maintenance data for four of the five students. Jada was unable to answer any of the problems correct during the first maintenance check (i.e., zero correct and six incorrect). The experimenter asked her to give directions following this performance in order to ensure she was aware of the task (i.e., she immediately jumped up to six correct and zero incorrect for the remainder of data points during maintenance). Abbie, Saddie, Jada, and Sasha all maintained one of their best performances (i.e., six correct, zero incorrect) following the TAPS/FBPC intervention. Theodore did not finish the intervention by the end of the school year and no maintenance data were collected. Overall, the maintenance data demonstrated that the students applied the skills learned during the intervention even after the instruction stopped. The maintenance data support the proposition of behavioral fluency that once a behavior reaches a frequency standard behavior persists for long periods of time after practice has terminated (Binder, 1996; Kubina, 2005; Kubina & Yurich, 2012; Kubina, Amato, Schwilk, & Therrien, 2008).

The TAPS/FBPC demonstrated a promising effect on problem solving when applied to science content. The intervention was also seen as beneficial by students who enjoyed the intervention and teachers who thought it was worthwhile. The study ran for over 100 days, allowing enough time for students to form an opinion of the frequently used problem-solving method. Despite the length of time, students liked using the techniques that they learned and would want to continue its use in the future.

### Future Directions and Limitations

Talk aloud problem-solving frequency building to a performance criterion holds promise in content area instruction. Additional exploration of the TAPS/FBPC is needed to determine more precise and efficient ways to allow for use in the classroom. The current study replicates Dembek and Kubina (in press), but additional data are needed to examine how the intervention can be used with a variety of populations and students. The experimenter required all students to meet the exit criterion (16 talk alouds about cues) prior to leaving the intervention phase, but in future replications more exploration should be focused on evaluating varied performance standards. Because performance standards are well-researched criteria that indicate a fluent performance, fluent levels are under scrutiny until further research is available.

The current study also displayed moderate variability in some data. Because of the nature of placing sentences in logical order, the answers directly affect one another. In other words, if a student gets the first sentence incorrect, they will automatically get another sentence incorrect that should be in that order (e.g., students says the order is 1, 3, 4, 2 but the order should be 2, 3, 4, 1). The experimenter corrected some of this variability with different scoring procedures, but there is no way to completely eliminate variability in this type of problem. Along with variability of data, transfer of talking aloud to writing answers should be addressed in future studies (i.e., seeing the prompts and saying the answer compared to seeing the prompts and writing the answer).

The current study explored TAPS/FBPC for students with disabilities and is the first to explore the use of the TAPS package of instruction for students diagnosed with a disability. Future studies should have generalization probes implemented frequently to determine any transfer in other content. Future studies might also explore the intervention with student pairs (i.e., one student as listener and the other as problem solver) rather than only the teacher serving in the role of expert. The study adds to the current literature working with older students in more advanced science course, a girl with autism, and a student-at-risk for disabilities. Replications are necessary to determine the effect on different ages, ability levels, and settings. Along with replications, more field tested probes would be beneficial for a more random selection of instruction materials. JEBPS

## REFERENCES

Bakken, J. P., Mastropieri, M. A., & Scruggs, T. E. (1997). Reading comprehension of expository science material and students with learning disabilities: A comparison of strategies. *Journal of Special Education, 31*, 300–324.

Banks, J. A., Beyer, B. K., Contreras, G., Ladson-Billings, G., McFarland, M. A., & Parker, W. C. (2001). *Regions: Adventures in time and place*. New York, NY: Glencoe/McGraw-Hill School Publishing Company.

Baumann, J. F., Seifert-Kessell, N., & Jones, L. A. (1992). Effect of think-aloud instruction on elementary students' comprehension monitoring abilities. *Journal of Reading Behavior, 24,* 143–172.

Bereiter, C., & Bird, M. (1985). Use of thinking aloud in identification and teaching of reading comprehension strategies. *Cognitive and Instruction, 2*, 131–156.

Binder, C. (1996). Behavioral fluency: Evolution of a new paradigm. *The Behavior Analyst*, *19*(2), 163–197.

Brigham, F. J., Scruggs, T. E., & Mastropieri, M. A. (2011). Science education for students with learning disabilities. *Learning Disabilities Research & Practice, 26*(4), 223–232.

Cooney, T., Cummins, J., Flood, J., Foots, B. K., Goldston, M. J., Key, S. G., & Weinber, S. (2006). *Scott Foresman science*. Glenview, IL: Pearson Education.

Cooper, J. O., Heron, T. E., & Heward, W. L. (2007). *Applied behavior analysis* (2nd ed.). Upper Saddle River, NJ: Pearson Prentice Hall.

Culler, E. D. (2010). *Effects of practicing passage retell to a fluency criterion to increase summarization*. (Doctoral dissertation). Retrieved from https://etda.libraries.psu.edu

Dawson-Boyd, C., Berkin, C., Geiger, R., Kracht, J. B., Ooka Pang, V., Risinger, C. F., … Meszaros, B. (2006). *Scott Foresman social studies: Regions*. Boston, MA: Prentice Hall School Division.

Dembek, G.A., & Kubina, R. M. (in press). The effect of talk aloud problem solving and frequency building to a performance standard with a student at-risk for reading disabilities: A case study. *Reading Improvement*.

Duschl, R., Schweingruber, H., & Shouse, A. (Eds.). (2007). *Taking science to school: Learning and teaching science in grades K-8*. Washington, DC: National Academies Press.

Fuchs, D., Fuchs, L. S., Mathes, P. G., & Lipsey, M. W. (2000). Reading differences between underachievers with and without learning disabilities: A meta-analysis. In R. Gersten, E. Schiller, & S. Vaughn (Eds.), *Research syntheses in special education* (pp. 81–104). Mahwah, NJ: Erlbaum.

Glass, A. R. (1992). *The effects of thinking aloud pair problem solving on technology education students' thinking processes, procedures, and problem solution*. (Doctoral dissertation, University of Minnesota). Retrieved from ProQuest Dissertations & Theses database. (UMI No. 303979616).

Hacket, J. K., Moyer, R. H., Vasquez, J., Terferi, M., Zike, D., LeRoy, K., … Bank Street College of Education. (2008). *Science: A closer look*. New York, NY: Macmillan/McGraw Hill.

Hasbrouck, J., & Tindal, G. A. (2006). Oral reading fluency norms: A valuable assessment tool for reading teachers. *The Reading Teacher*, *59*(7), 636–644.

Heil, D., Allen, M., Cooney, T., Matamoros, A., Perry, M., & Slesnick, I. (1994). *Scott foresman science: Discover the wonder*. Glenveiw, IL: Pearson Scott Foresman.

Holzer, S. M., & Andruet, R. H. (2000). Experiential learning in mechanics with multimedia. *International Journal of Engineering Education, 16*(5), 372–384.

Jeon, K., Huffman, D., & Noh, T. (2005). The effects of thinking aloud pair problem solving on high school students' chemistry problem-solving performance and verbal interaction. *Journal of Chemical Education, 82*(10), 1558–1564.

Johnson, S. D., & Chung, S.-P. (1999). The effect of thinking aloud pair problem solving on the troubleshooting ability of aviation technician students. *Journal of Industrial Teacher Education, 37*(1), 1–18.

Johnston, J. M., & Pennypacker, H. S. (2009). *Strategies and tactics of behavioral research*. New York: Routledge.

Kani, N. H. A., & Shahrill, M. (2015). Applying the thinking aloud pair problem solving strategy in mathematics lessons. *Asian Journal of Management Sciences & Education, 4*(2), 20–28. Retrieved June, 2015, from http://www.researchgate.net/publication/275643101

Kennedy, C. H. (2005). *Single-case designs for educational research*. Boston: Allyn & Bacon.

Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2013). Single-case intervention research design standards. *Remedial and Special Education, 34*, 26–38.

Kubina, R. M. (2005). The relationship between fluency, rate building and practice: A response to Doughty, Chase, and O'Shields. *The Behavior Analyst, 28*, 73–76.

Kubina, R. M., Amato, J., Schwilk, C. L., & Therrien, W. J. (2008). Comparing performance standards on the retention of words read correctly per minute. *Journal of Behavioral Education, 17*, 328–338.

Kubina, R. M., & Yurich, K. K. L. (2012). *The precision teaching book*. Lemont, PA: Greatness Achieved.

Mason, L. H., & Hedin, L. R. (2011). Reading science text: Challenges for students with learning disabilities and considerations for teachers. *Learning Disabilities Research & Practice*, *26*(4), 214–222.

National Center for Education Statistics. (2015). *The Nation's Report Card: Science 2011* (NCES 2012–465). Washington, DC: Institute of Education Sciences, U.S. Department of Education.

The Nation's Report Card. (2015). National Achievement-Level Results: Grade 8-Science. Washington, DC: Institute of Education Sciences, U.S. Department of Education. Retrieved from https://www.nationsreportcard.gov/science_2015/#acl?grade=8 [July 2018]

Pate, M. L., & Miller, G. (2011). Effects of think-aloud pair problem solving on secondary-level students' performance in career and technical education courses. *Journal of Agricultural Education*, *52*(1), 120–131.

Pate, M. L., Wardlow, G. W., & Johnson, D. M. (2004). Effects of thinking aloud pair problem solving on the troubleshooting performance of undergraduate agriculture students in a power technology course. *Journal of Agriculture Education, 45*(4), 1–11.

Pate, M. L., & Young, C. (2014). Compact power equipment troubleshooting training: Formative assessment using think-aloud pair problem solving. *NACTA Journal, 58*(3), 256–261. Retrieved from http://search.proquest.com/docview/1613183435?accountid=7286

Pennypacker, H. S., Gutierrez, A., & Lindsley, O. R. (2003). *Handbook of the standard celeration chart*. Cambridge, MA: Cambridge Center for Behavioral Studies.

Pestel, B. C. (1993). Teaching problem solving without modeling through "thinking aloud pair problem solving." *Science Education, 77*(1), 83–94.

Robbins, J. K. (2011). Problem solving, reasoning, and analytical thinking in a classroom environment. *The Behavior Analyst Today,* 12, 42–50.

Saenz, L. M., & Fuchs, L. S. (2002). Examining the reading difficulty of secondary students with learning disabilities: Expository versus narrative text. *Remedial and Special Education*, *23*, 31–41.

*Sciencesaurus: A student handbook*. (2002). Wilmington, MA: Great Source Education Group.

Tingle, J. B., & Good, R. (1990). Effects of cooperative grouping on stoichiometric problem solving in high school chemistry. *Journal of Research in Science Teaching, 27*(7), 641–683.

U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics. (2015). NAEP data explorer. Retrieved from https://www.nationsreportcard.gov/science_2015/#acl?grade=8 [July 2018]

What Works Clearinghouse. (2011). WWC procedures and standards handbook. Retrieved from http://ies.edu.gov/ncee/wwc/DocumentSum.aspx?sid=19

Whimbey, A. (1995). *Mastering reading through reasoning*. Raleigh, NC: Innovative Sciences.

Whimbey, A., & Lochhead, J. (1999). *Problem solving and comprehension*. Lawrence Erlbaum. Mahwah, NJ: Lawrence Erlbaum.

White, O. R. (1986). Precision teaching – precision learning. *Exceptional Children, 52*, 522–534.

Witcoski, G. A., & Kubina, R. M. (2012). *Thinking aloud and reading comprehension: A review of intervention research*. Unpublished manuscript, Department of Special Education, The Pennsylvania State University, University Park, PA.

Wong, B. Y., & Wilson, M. (1984). Investigating awareness of and teaching passage organization in learning disabled children. *Journal of Learning Disabilities*, *17*(8), 477–482.

**Ginny A. Dembek**, PhD, Department of Childhood, Bilingual, and Special Education, Brooklyn College CUNY. Dr. Dembek is an assistant professor in the School of Education. Her areas of research interest include literacy interventions, measurably effective classroom practices, and precision teaching.

**Richard M. Kubina** Jr., PhD, Special Education Program, The Pennsylvania State University. Dr. Kubina is a professor in the College of Education. His areas of interest are precision teaching, applied behavior analysis, single case research design, and technology.