Classroom Assessment for Language Teaching

Classroom Assessment for Language Teaching:

Challenges, Choices, and Consequences

Edited by

Eddy White and Tahnee Bucher

Classroom Assessment for Language Teaching:
Challenges, Choices, and Consequences

Edited by Eddy White and Tahnee Bucher

This book first published 2020

TABLE OF CONTENTS

Section 4 – Assessing Grammar

Section 5 – Assessment in IEPs and Pathway Programs

Section 6 – Teacher Assessment Literacy

PREFACE

It is now commonly recognized that assessment literacy, the understanding and know-how about sound assessment principles and practices, is a valuable and essential skill for both teachers in mainstream education as well as in language teaching. It is necessary for teachers' long-term professional well-being, in addition to benefitting the students and institutions where they work. As such, a central concern for teachers and the English language teaching profession is the enhancement and promotion of teachers' assessment competence. While the importance of such assessment competence is widely acknowledged, it is also recognized (and bemoaned) that competence in assessment to both promote and measure student learning is not only lacking in practicing teachers, but is also neglected in the literature and conferences associated with English language teaching.

Reports from numerous surveys and studies show that teachers are sorely lacking in their knowledge and skills related to the assessment of student learning, and while the reasons for such shortcomings are debated, there is no denying its existence. In the most recent edition of their *Teacher-made Assessments* book, Gareis and Grant (2015) assert the following: "*Classroom-based Assessment is an extraordinarily important component of teaching and learning that has long received insufficient attention*" (p. 174). Much of the research and publications focus on what has been termed the 'testing industry'—the large-scale commercial instruments that typically are used for proficiency testing purposes. While playing an important function and role in language education, the associated research and publications can often seem very far removed from the classroom, and the challenges that teachers face in assessing and grading their students' language knowledge and skills. Clearly there is much room for publishing and promoting an increased awareness of assessment as it plays out for learners and teachers in the language classroom.

This book is clearly targeted at helping address this issue of 'insufficient attention' being paid to assessment at the classroom level, and helping promote classroom assessment literacy in our profession. It seems reasonable to assume that all would agree with the assertion that the more assessment literate that teachers and administrators at a language teaching institution are, the better it will be for all stakeholders—for students (of primary importance), teachers, administrators, programs, and the institution

itself. It is hoped that this book will provide a valuable resource for teachers wrestling with the challenges and opportunities associated with classroom assessment.

Since the 1970s there has been a group of educators and researchers who have argued that the key to being good teachers lies in the *decisions* that teachers make. Good classroom assessment lies, therefore, at the core of good decision-making. Key components of this collected edition are the choices that classroom teachers (and administrators) make in responding to challenges they face related to classroom assessment, and the consequences (both positive and negative, anticipated and unexpected) that result from their decision-making.

This book also has a key focus on practitioner research, whereby the teacher has the dual role of both practitioner and researcher in sharing the voices and stories arising from the classroom context. Focused on a key challenge or question(s), the teacher-researcher's reports of their classroom investigations use relevant data collected, as well as instruments and procedures used, and the results of their decisions to provide an analysis and conclusions to what they discovered. Investigating and reporting, in a detailed and systematic manner, on their response to the classroom assessment challenges faced places the classroom setting and participants at the heart of such practitioner research.

A compilation of such research reports provides a valuable collection of classroom assessment reports for teachers and programs in various other institutions, as well as obviously promoting the assessment literacy of the researchers/contributors to the special edition. Teachers who use assessments need to understand what the results of assessments mean, and also, what assessments do to those involved in the assessment process; in other words, teachers' assessment-related decisions (and assessment literacy) have both evidential and consequential aspects. Such evidential and consequential aspects are critical components of the investigations making up this book.

A better understanding of the qualities of good assessments and of processes and techniques for improving this competency should help practicing teachers and other educators to make more effective use of assessment in their professional lives. Teachers themselves lie at the heart of this endeavor. Ultimately, the value of assessments depends on the people involved in designing, building, administering and scoring them, and those who interpret the results. Good practice in language assessment, finding the most appropriate tools for each purpose, requires commitment and expertise. It is hoped that this collection of investigations focused on such classroom assessment practices and competencies, and the students and

teachers involved, can promote such assessment literacy and best practices for the good of all concerned.

Chapter Focus/Organization

This collected edition of ten chapters follows a specific format, an organizational structure that readers will easily note, and indeed is made clear in the book title.

The book focuses on the experiences and voices of teachers in language classrooms around the world. We invited teachers to write about the assessment challenges they face and related investigations they conduct in dealing with those challenges. We reminded potential authors that language teachers everywhere deal with similar issues and challenges (e.g., how to set up a course grading framework, how to best assess speaking skills, how to deal with plagiarism and cheating), and there is much to learn from each other. This collection of chapters not only promotes assessment practices among professionals, but also provides multiple perspectives from the field, from the point of view of the teachers themselves!

Specifically, the chapters are loosely organized as reflected in the book title (i.e., challenges, choices, and consequences). Authors were asked to address the following questions:

- Describe a particular challenge you faced in assessing/grading students in the language classroom.
- What choice(s) did you make in dealing with this challenge?
- What were the consequences of your decision-making (or the anticipated consequences)?

As readers will note, a key requirement for authors submitting manuscripts was that their writing be firmly anchored to the published literature and best practices related to their particular area (e.g., assessing essays).

The 10 chapters are compiled into six sections as follows: Assessment and Emotions, Assessment in Conversation Classes, Assessing Writing, Assessing Grammar, Assessment in IEPs and Pathway Programs, and Teacher Assessment Literacy.

We are very pleased to be able to collect this series of voices, stories, and investigations from various countries, perspectives, and aspects of classroom assessment. We hope it makes a useful contribution to the growing field of classroom-based assessment in language teaching. In particular, we hope that it provides a resource for teachers and administrators

dealing with similar assessment challenges and choices with their programs, courses, and language students.

Eddy White, PhD
Tahnee Bucher

ACKNOWLEDGMENTS

SECTION 1 –

ASSESSMENT AND EMOTIONS

CHAPTER 1

TESTING ENVIRONMENT, EMOTIONS, AND STUDENT PERFORMANCE: CLASSROOM AND PROGRAM WIDE

SHELLY STODDARD AND ERIC MOORE
VIRGINIA POLYTECHNIC INSTITUTE AND STATE UNIVERSITY

Abstract

The testing environment is a critical aspect of student performance. Students who experience anxiety and other negative emotions when faced with an assessment generally perform at a lower level than their peers who feel confident and secure in their abilities. This chapter explores the relationship of emotional barriers on student achievement and provides strategies to mitigate them. Through a comprehensive literature review, the authors found evidence that test anxiety, previous assessment experiences, cultural background and self-efficacy beliefs, and the relationship between the student and instructor impact performance on assessments. The authors recommend several strategies to overcome these negative hurdles in order to improve performance and achievement as well as ensure reliability in classroom and program outcomes. Classroom strategies include building schema for testing vocabulary and protocols, using formative assessment to influence instruction, implementing constructive feedback and metacognitive habits, and creating strong relationships between the teacher and student. Program wide strategies focus on teacher training, standardization, and the use of authentic assessments. In addition to sharing strategies, the authors describe the challenges faced by administrators and teachers when implementing changes to assessment routines and the success achieved by students after these new strategies were realized.

Introduction

The instructor looks across her classroom and observes her students awaiting the assessment she promised. Some students squirm nervously in their hard-plastic chairs while others have red blotches creeping up their neck and into their cheeks. Some students tap their pencils as if counting down, while a few seem to be silently praying to their spiritual guide. Still others quietly wait and whisper to their neighbor seemingly unaware that this task holds more weight than general classwork. This scene may sound familiar and it illustrates the variety of responses students can have when faced with assessments. Emotional responses to the words *assessment* or *test* exist on a continuum. The testing environment and process, relationship between the student and teacher, and student background shape how these emotional hindrances affect achievement for some students and program outcomes.

Emotional responses present in a variety of ways. The most commonly observed response is test anxiety, which is categorized into two types: trait and state (Speilberger & Vagg, 1995). Trait anxiety is defined as a more permanent and pervasive emotional tendency based on personality and disposition, whereas state anxiety is experienced when a person is faced with a specific task at a particular time (Grant, 2016). In the classroom setting, students who exhibit anxious characteristics when faced with assessment situations are generally undergoing the latter type of anxiety. All types of students can experience some degree of test anxiety, and English language learners are no exception. In fact, those learners studying in a second language learning environment are especially susceptible to anxiety due to additional stressors they face including recent immigration, acclimation to a new culture, and introduction to the American education system (Grant, 2016).

Another factor influencing emotional responses to assessment is individual perception. Both teachers and students may assign a greater value to a task labeled *assessment* and therefore respond to the situation based on one's personal perspective. This can lead to students or teachers attributing more or less value to an assigned task and thereby influence its results with a greater negative emotional response. Jung, Wranke, Hamburger, and Knauff (2014) purported that a person's emotional state affects their performance on a task. Their analysis also suggested that one's cognitive ability is impacted by their emotional condition, and this can influence motivation and attention to task.

Moreover, emotional reactions to testing are further influenced by a student's cultural background and previous testing experiences in their

native country. For example, our largest population of students is from China, and they have been rigorously groomed and prepared for most of their primary and secondary school careers to take a formidable assessment called the *Gaokao*, which determines their future university experience and job opportunities (Muthanna & Sang, 2015). This assessment is associated with long hours of preparation, high pressure and stress, and recently a rise in suicide among young people (Davey, Lian, & Higgins, 2007). According to Muthanna and Sang (2015), the Gaokao exam is the most pressure laden exam worldwide and surpasses exams like the SAT in its competitiveness and long-term outcomes. Students who relate this type of pressure and intensity to testing and examinations might transfer this emotional response to all testing situations.

Another predominant group served at our university are students from Saudi Arabia. According to Alrabai (2018), a key indicator of success is one's self efficacy beliefs. Bandura (1997) purported that self-efficacy includes one's ability to organize and accomplish required tasks as well as one's beliefs in learning new things given the skills he or she possesses in any given situation. Self-efficacy is based on three types of experience: mastery, vicarious, and verbal. These experiences are often unavailable to Saudi students or result in negative self-talk resulting in lower achievement overall (Alrabai, 2018). Furthermore, "a stressful learning environment fueled by nervousness, tension, low confidence and motivation, and negative attitudes as seen in the Saudi EFL context" (Alrabai, 2018, p. 1358) contributes to the development of low self-efficacy and therefore reduced achievement in the classroom setting. Students who view themselves as unable to learn or progress in the language classroom may self-sabotage their efforts in a testing environment regardless of what was taught or learned.

At our large university located in the southeastern United States, instructors and test administrators observed students performing below the expected benchmarks in multiple settings, including summative and formative classroom tests, whole institution assessments, and proficiency assessments for matriculated international students. At our university, our program is part of the Outreach and International Affairs Department and we currently have three operating areas within our program: the Intensive English Program (IEP), the Pathways Program (AdvantageVT), and Global Initiatives. Class sizes range from as few as five to as many as cohorts of thirty students. In the IEP, we serve all levels of students from entry level to the advanced level when students are ready for university entrance. In the AVT Program, students enter with an English proficiency score of sixty on the Test of English as a Foreign Language (TOEFL iBT) and matriculate

over the course of two semesters while continuing their language development. Learning outcomes for students range in degree of difficulty with a focus on skill and vocabulary development in all language areas including speaking, listening, reading, and writing in the IEP. The AVT program focuses more on twenty first century skills and content area themes to fully prepare students for both academic and language competence once they are fully matriculated.

Assessments drive most educational decisions within our program. Some assessments like those used in the classroom are controlled and created by instructors while the program wide assessments are created in committee and approved by the administrative team. Therefore, when the testing environment creates undue stress and anxiety resulting in lower performance, these educational decisions may be unreliable resulting in incorrect placement, missed opportunities, reduced self-esteem, and lack of progression (Gebril, 2018). This was not only adverse for our students but also for our program and institution. Consequently, we believed creating testing environments that overcome these emotional hurdles might significantly increase student performance and by extension create better programming. At our institution, we implemented a variety of methods to combat this issue. These targeted strategies included building rapport, creating and using authentic tasks, pre-teaching testing protocols and processes, building schema, teaching testing vocabulary, soliciting student feedback, and implementing reflective strategies.

Challenges

Taking assessments to demonstrate proficiency and using assessment results to make decisions generate challenges for both students and educators. These challenges are twofold: one is based on the teacher's or administrator's role, while the other is focused on the student's perception and response. Listed below are the assessment-related issues that our institution has faced.

- Student emotional responses to a variety of assessment situations
- Impacts of perception, background, and values on the role of assessment
- Creating testing environments that reduce negative emotional responses
- Ensuring that standardization exists within program wide assessment

Student Emotional Responses

"You have a test tomorrow" is such a simple statement, but for some students it has a profound and sometimes debilitating effect. These negative effects can result in behavioral, physical, and affective symptoms and is often referred to as test anxiety. Like other emotional responses, test anxiety occurs on a range with those severely impacted to those with just a moderate response. Test anxiety is estimated to intensely affect as many as 16% to 20% of students and moderately affect as many as 18%, resulting in students 'freezing or blanking out' when faced with an assessment. "High test anxiety reduces working memory, confuses reasoning, increases mistakes, and lowers test scores. Students with high anxiety perform around twelve percentile points below their low anxiety peers" (American Test Anxiety Association, n.d., para. 2).

Zeidner and Matthews (as cited by Huntley, Young, Temple, Longworth, Smith, Jha, & Fisher, 2019, p. 4), stated that test anxiety encompasses the "phenomenological, affective, and behavioural responses that accompany concern about the possible negative consequences of poor performance in an examination or other performance evaluative situation." Speilberger and Vagg (1995) argued that educational and professional opportunities are often decreased because the negative effects of test anxiety result in a minimized view of an individual's actual ability and skill. In addition, Steinberg and Horwitz (1986) noted that anxiety, specifically in second language learners, has both an impact on discrete skill evaluation, idea expansion, fluency, and grammatical structure in speaking and writing. Yan and Horwitz (2008) asserted that there are several factors which influence language learner anxiety, including "motivation, cultural factors, the students' own coping skills, attention, self-concept, beliefs about language learning, and the specific teaching methodology the student experiences" (p. 152). Moreover, Young suggested that there are six categories of language learner anxiety which include "personal and interpersonal anxieties, learner beliefs about language learning, instructor beliefs about language teaching, instructor-learner interactions, classroom procedures, and language testing" (as cited by Yan & Horwitz, 2008, p. 152). Based on this literature review, test anxiety is an emotional barrier which may reduce a student's ability to perform and demonstrate mastery in a testing situation.

Although test anxiety is a commonly accepted response to assessment, it is not the only response. According to Thomas, Cassady, and Heller (2017), "student performance can be adversely impacted by stressors within and beyond the academic setting" (p. 40). These internal stressors may include the difficulty of a task, anxiety due to subject matter, and perceived academic burden, while external stressors may include financial issues,

personal needs, and familial responsibilities. Likewise, student social emotional temperaments influence their ability to manage their environment, specifically their response to a negatively perceived task (Thomas et al., 2017). Chin, Williams, Taylor, and Harvey (2017) suggested that cognitive factors such as "worries about potential failure and its consequences, feeling unprepared for the test, and having self-oriented doubts over one's own capabilities" (p. 1) are related to decreased performance on assessments. Additional causes of lower performance include working memory impairment resulting in less available resources when taking the exam, student academic goals for success and achievement, procrastination, ineffective coping skills, and a lack of academic resilience (Chin et al., 2017).

Speilberger and Vagg (1995) reviewed several student case studies and found that students who were well adjusted and functioned effectively in most aspects of their emotional lives were impacted by anxiety on specific testing items. Conversely, students who had emotional problems that interrupted their lives often had test anxiety, but it was just one expression of their overall generalized emotional state. One case study focused on Jane, a well-adjusted college student, who appeared to be emotionally mature. She tended to perseverate over her test performance while taking the exam, which led to limited impacts on test performance. Another case study subject named Jim was dealing with a personal crisis which led to increased stress over his performance because he felt he had to prove his value as well as his overall aptitude as a student. Ted, a third case study participant, was a systematic and responsible student with perfectionist tendencies. His beliefs about himself and his general anxiety impacted his overall performance on assessments. All of these case studies demonstrate that one's personal beliefs about self and one's environment, past experiences, and emotional stability influence emotional responses felt in the assessment setting.

Similarly, our students presented with a variety of responses when they were asked to perform on an assessment. These responses included:

- Negative self-talk, increased worry, and indifference:
  - Students verbally or mentally reprimanded themselves for perceived performance errors and believed they were incapable of achievement. This resulted in a completion of the task in a hurried manner as though it was irrelevant. Students also demonstrated increased worry and concern through their body language, appearance, and actions.

- Paralysis and inability to perform the task:
  - Students freeze and are incapable of completing the task.
- A female Egyptian student was in our mid-level speaking and listening class. In class, she could easily talk with her peers and could present information to the whole group in an effective way using level appropriate English vocabulary and grammar structure. However, when asked to complete the end of term Speaking Assessment, she was unable to speak and was reduced to stuttering, unintelligible language, and physiological responses like headache and shaking. When asked, she explained that she was so nervous that she just could not perform (M., Abdelmaksoud, personal communication, April 13, 2017).
- Cheating:
  - Students are so afraid of poor performance that they resort to cheating
- A Chinese male student in a mid-level reading and writing class was caught using a slip of paper taped to his hat with vocabulary words and their definitions while taking a test. When asked about his actions, he claimed that he was so afraid of a low grade that he needed support to complete the task even though his teacher had asked him what the words meant in an informal setting and he was able to do so without error.
- Lower performance compared to in-class tasks:
  - Students perform at a proficient level on 'non-test' tasks, but when asked to complete an assessment, increased anxiety leads to lower performance.
- A female Saudi Arabian student in an intermediate level writing class was asked to complete a timed writing test. She was asked to write about a topic that she was not able to prepare for in advance, using the same essay format that she had used in class. She performed poorly compared to her in-class writing samples. When asked, she explained that she felt frozen and as though her mind was blocked and could not think of ideas or words to express them.

Clearly, negative emotional responses like test anxiety and worry, personal beliefs about one's self, and ineffective coping mechanisms impact student performance. At our institution, we used these observations to make changes to our assessments and the testing environment. These choices were made to combat decreased achievement resulting from emotional reactions rather than actual educational deficits.

Impacts of Perception, Background, and Values

In the United States, education is driven by tests. Students are given standardized assessments throughout their educational career to measure both their performance against their peers and their knowledge of required subject matter. Test scores are used to rank students, elevate grade point averages, receive scholarships, and enroll in university. American students are accustomed to testing, yet they recognize that it is only one aspect of their education and ability. To enter university, extracurricular activities, volunteer experience, and recommendations help distinguish students for college entrance, not just test scores (Kowarski, 2018).

American educators who experienced this type of testing environment may view testing as necessary, but not all encompassing. Teachers may also expect students to prepare for exams and demonstrate their ability in testing situations because they attach a higher value to tasks labeled 'test' compared to those called 'homework'. Additionally, teachers with little experience with culturally responsive teaching practices may not be familiar with how other cultures view testing and its impacts on performance. Finally, teacher training in assessment is not standardized and varies widely among universities and programs. Crusan, Plakans, and Gebril (2016) suggested that teachers often feel unprepared, lack adequate training in the area of assessment literacy, and require more education in assessment development, scoring, and evaluation.

Students from around the world experience testing with different degrees of pressure and success. As China modernizes, increased numbers of Chinese students have sought education in the United States and other English-speaking countries. These students have been exposed to intensive testing using the state sponsored university entrance exam (Gaokao). This exam is responsible for students entering university or the workforce and dictates one's entire future. Increased pressure from parents and long hours of preparation are common for Chinese students. As a result, they may view all testing settings in a similar fashion (Davey, Lian, & Higgins, 2007). Additionally, Davey et al. (2007) suggested that students from China typically have a different set of educational skills as the Chinese educational frameworks differ from those in the West.

In Saudi Arabia, students typically have low self-efficacy resulting in decreased achievement in language learning. Branscombe and Baron (2016) divide self-efficacy into three main categories: self-regulatory, social, and academic. Self-regulatory efficacy includes one's ability to resist peer pressure and avoid situations where elevated risk is involved. Social efficacy includes one's ability to create and maintain interpersonal relationships, and academic efficacy is the ability to manage school work,

organize learning tasks, and successfully meet personal expectations. Saudi students must be taught that success is achieved through effort and discipline and is not an exclusive function of one's ability (Alrabai, 2018). These beliefs about self directly impact student performance and can result in lower achievement.

Educator knowledge of differing cultural experiences and how they influence performance has an effect on overall testing performance. Our institution found that our two major demographic student groups exhibited distinct testing behaviors and our methods changed to address performance in both groups.

The Testing Environment, Program Assessment Protocols, and Standardization

Desks lined up in rows, silence, and a teacher walking around the perimeter of the room watching with laser like focus is an image many students and educators imagine when the test environment is mentioned. Our challenge was that the testing environment did not promote student success. Strict adherence to scripts, required testing templates, and a lack of testing experience led to decreased achievement on formative and summative exams.

Some teachers felt that what was being tested on the summative exam did not match how it was taught. Some felt obliged to follow a script or set of directions that did not fit the group of students being assessed. Still others failed to pre-teach or practice the exam format resulting in students performing poorly simply because they had no prior knowledge of the technology or format being used. Additionally, some teachers were very strict in their adherence to guidelines while others opted to make adjustments, including unintentionally giving too little information, which resulted in students missing key directions and time limits thereby impacting performance. Lastly, some teachers were perceived as maternal or paternal while others were seen as friends and still others were viewed as authoritarian resulting in a variety of student behaviors that impacted the testing environment negatively.

Challenges: Conclusion

Emotional responses, individual perceptions, and the assessment setting and conventions altered student achievement at our institution. Student performance in class did not regularly align with the program wide assessments. Testing in one classroom to the next was varied in test delivery

as well as preparation. Specific student populations responded differently when faced with assessments, and test anxiety created achievement gaps between performance on class assignments and success on tests. In the next section, the choices we made to address these challenges are explained.

Choices

The choices we made at our institution can be categorized into two groups—program wide and classroom specific. Some choices were easily implemented and used immediately while others required pilots, revisions, training, and instructor feedback. Authentic Speaking Interviews and Process Based Timed Writing (PBTW) are currently used program wide in our Intensive English Program (IEP). Classroom assessments focus on both formative and summative assessments. Our institution made specific choices to overcome emotional barriers for both types as well as enhance the overall testing environment to promote learner achievement.

Authentic Speaking Interviews

Several years ago, in order to meet new United States Department of State regulations on proof of English language proficiency levels of visitors arriving on J-1 visas, we were tasked with designing and employing an assessment program for our university. It later became known as the J-1 Exchange Visitor English Language Proficiency Interview (ELPI). Our goal was to create an environment of well-being to enable test takers to relax and become comfortable speaking with us, which we considered the best way to get a true representation of a test taker's ability. This view is based upon best practice as defined by Bachman and Palmer (1996), who offered two important ideas related to this test in their language testing philosophy: designing tests in order to encourage and assist test takers in performing as best they can, and humanizing the test process.

We felt that we would get more authentic language and a truer, more representative sample of language if we established a rapport with test takers. Rapport can be described as a "personal connection and an enjoyable interaction" (Webb & Barrett, 2014, p. 11). A strong rapport is likely to decrease test anxiety (Neil & Christensen, 2009). Gremler and Gwinner's work on rapport (2000) described in Webb and Barrett's (2014) research, lists five tactics employed to build rapport—common grounding behavior, connecting behavior, courteous behavior, information sharing behavior, and uncommonly attentive behavior.

Common grounding behavior is used to express common interests, such as a discussion of NBA professional basketball players with a basketball fan. Connecting behavior is employed to develop an interpersonal bond. For example, the test proctor shares a short personal story with a test taker about living in a different country. Courteous behavior is easily engaged by being considerate, kind, and empathetic. Information sharing behavior can be demonstrated by using clear communication of relevant test information, such as describing to a test taker that grammar usage will not be scored on a test of pronunciation. Uncommonly attentive behavior can be brought to bear by going above and beyond to make sure all attention is focused on the test taker rather than on scoring rubrics or focusing on the clock. Brown (1994) described the importance of showing interest in the students and issuing an invitation to students to express how they feel. ELPI administrators accomplished this by asking focused follow-up questions based on the test taker's interests and responses.

We viewed building rapport as one of the two cornerstones of the ELPI. The other was the employment of task-based assessment items. Since we wished to learn as much as we could in a short period about our test taker's English language skill, we felt that task-based items would serve us well. Norris (2016, p. 241) pointed out that "task-based performances can reveal multiple aspects of language ability and/or development within a single instance (e.g., accuracy, complexity, and fluency; content knowledge; procedural knowledge; expertise; and pragmatic sensibility." Appendix A exhibits examples of questions posed on the ELPI.

Test takers shared positive feedback through survey responses following the interviews we conducted (see Appendices B and C). Based on the effectiveness our rapport techniques were generating, we believed this same choice could be applied to the IEP program's speaking exams. At the end of the fall semester of 2016, we piloted speaking final exams in our IEP classes. The following list shows the procedures course instructors were trained to use on the exams (Fulcher, 2014; Luoma, 2004; Underhill, 2004).

- Create a comfortable environment by using students' names and offering a welcome greeting at the beginning of the test.
- Remind students about the test format and scoring.
- Begin the interview with short, easier to answer questions to build confidence and comfort.
- Employ follow-up questions in areas of obvious interest in order to obtain longer speaking samples.
- Take leave in a positive, authentic manner at the completion of the assessment.

Process-Based Timed Writing

Our IEP had been assessing student learning outcomes in writing through timed writing tests since the early years of the program. Standardization and accountability were lacking as each instructor created her own test and scored the test herself. This was problematic for several reasons. A student in Instructor X's class could have a higher chance of achieving a passing score than in Instructor Y's course at the same level despite the same amount of skill, production, and effort. In addition, there was a large amount of variability and a reliance on outdated testing and assessment practices in teacher-based assessments (Davison & Leung, 2009). Finally, accreditation agencies have also focused on program-wide standardization as an area of high importance for IEP programs. The administrative team felt that changes were necessary to ensure standardization and accountability existed across all aspects of the program.

In 2015, the Testing and Assessment Coordinator (TAC) and one of the instructional level leaders attended a TESOL international conference session that presented information on an integrated process-based timed writing (PBTW) exam (David, 2015). The presentation compared a traditional timed writing exam to the PBTW. We felt that there would be several advantages to adopting the exam. Advantages included increased familiarity (He & Shi, 2008), additional planning time prior to writing (Worden, 2009), and greater schema for students to draw on in order to successfully write about a topic (Esmaeili, 2002).

This type of exam had two disadvantages. Instead of taking one fifty-minute class period, it required two periods. This required teachers to adjust their pacing guides to reflect time dedicated to two class periods for testing plus time needed to train students on the new process. In addition, it was not as easy to administer as a traditional timed writing exam. Previously, the exam was simply handed out, read aloud, and then students had fifty minutes to complete it. With this new process, it would require that teachers instruct students prior to the exam on how the process would work, as well as follow a new process. This process included watching background building videos, reading related texts, having a class discussion and group discussion, and then reading the writing prompt. Teachers had to accept the new procedure in order to ensure that consistency was present in all testing rooms during the final timed writing. Lastly, it was important that students also embraced the new process and saw the benefits of building background information on a topic so that the writing task only measured writing.

Both attendees felt that the PBTW, if adopted, would offer our organization an opportunity to transition to a final written exam based upon the type of tasks required in university content area courses. Adopting the PBTW

would also allow our program a chance to reestablish standardization and accountability through training and observation.

The TAC designed a series of steps to follow both before and on the test day. These steps are exhibited in Appendix D. Instructions were given to instructors to familiarize both themselves and their students with the new process early in the term. Instructors were encouraged to practice the format throughout the term as students would benefit if they understood the test format and scoring method ahead of the actual assessment (Educational Testing Service, 2013).

The testing process began with the instructor leading a general discussion of the topic during the first hour of class. Following this, instructors showed a video(s) to the students. Next, a reading passage(s) was distributed to the students. Students were allowed and encouraged to take notes during any part of the first half of class. The remainder of the first half of the full two-hour testing window was spent in small group discussions with the instructor circulating around the room answering questions or contributing briefly to discussions. Although students had not seen the prompt yet, they typically had a general idea of what to expect. After a ten-minute break, the students returned to the room. The student response sheet was handed out to the students. They were asked to read the prompt and could use the notes taken in the first half of the exam to plan and write their response to the prompt during the second half of class which was fifty minutes in length.

Classroom Assessment

Formative assessment

Formative assessment is defined as "the process of monitoring student knowledge and understanding instruction to ensure maximal student growth" (Noyce & Hickey, 2011, p. 1). Instructors who use formative assessment successfully and continuously ask probing questions, analyze student responses, listen while students work together, create and review discrete tasks and quizzes associated with a skill. Next, teachers make minor adjustments based on this real-time analysis to redirect or alter the instructional plan for individual student needs to ensure the greatest impact on student growth and performance. The most adept instructors are almost able to do this instinctively while others need more explicit instructions on how to collect and analyze data to adjust instruction (Noyce & Hickey, 2011).

Informal formative assessments

Informal formative assessment is generally completed during the instructional cycle and used to inform a teacher's instructional practice to directly impact student growth and learning (Conderman & Hedin, 2012). To reduce negative emotional responses and create testing environments that directly combat the challenges we saw in our institution, some teachers implemented systematic informal formative assessments. These assessments included warm-ups, technological response system use, anticipation sets, pre-tests, graphic organizers, and exit slips. These types of assessments gave the instructors a quickly accessed sample set of what her students knew ahead of a lesson, during a lesson, or after a specific set of instructional outcomes were taught. Instructors shared that they preferred this method of assessment as it could be done in real-time and lessons could be adjusted in the moment to effectively address student deficits or extend a lesson for those who were exceeding the benchmark (C. Capone, personal communication, February 17, 2019). Table 1 highlights how these assessments were used in our institution's classrooms.

Table 1: *Informal Formative Assessment Examples*

Warm Ups	As students enter, they complete three gap fills on new vocabulary and three gap fills on vocabulary taught from the previous day.
Anticipation Guides	At the beginning of the unit on sustainability, students are asked to complete an evaluation of their school building and home.
Response Systems (technology or dry erase boards)	Students have studied a specific grammar structure in class. When given gap fills as a whole group, they respond by writing their answer on a dry erase board or using a smart phone application like Kahoot.
Pretests	Students are given a pretest on content they are going to be taught. Items may include a concept that is new.
Graphic Organizers	As new information is taught, students use graphic organizers to organize the new information. These include circle maps, concepts maps, tree maps, flow maps, KWL charts, the frayer model, defining format map, and t-charts.

Exit Slips	Students are given an index card at the conclusion of a lesson or class. On the card, they are asked to write sentences using 3 new vocabulary words and relative clauses. (This is based on what was taught during the lesson.)

Warm-ups are gap-fills, error correction, or open-ended questions that allow students to demonstrate what they know based on a previous lesson, or show a teacher what a student knows

prior to a new lesson. Like warm-ups, pretests can be used to assess what students know prior to a unit of study giving instructors valuable information for planning purposes. Anticipation guides are especially engaging as they captivate student attention and utilize higher order thinking skills. Response systems are an effective tool for instructors to assess current knowledge and then determine whether to reteach a concept or not. They also allow teachers to give corrective feedback when cards, dry erase boards, or devices are being used. Moreover, teachers who use technological systems excite and motivate students by integrating smartphone or interactive response system use into a lesson. Technological response platforms like Kahoot also allow teachers to download spreadsheets for analysis and data collection. Graphic organizers can be used at any time during a unit of study, but are most commonly used during instruction for the purposes of organizing information. These tools give students an easy way to organize new content so that it can be more easily assimilated. Additionally, teachers can review the organizers to check what students are comprehending and where gaps might exist. Finally, exit slips are used to quickly assess what students understand from a lesson taught that day. They can be a simple index card with three questions or directions to write three examples of a concept. Teachers can also create exit slips that incorporate the 3-2-1 strategy where students write three things they learned, two things they still want to know, and one clarifying question. These are especially beneficial to instructional planning as teachers can quickly identify patterns and note what to reteach and when to move on to the next concept (Conderman & Hedin, 2012).

At our institution, instructors indicated that more detail and information was understood from formative assessment on student ability at regular intervals. For example, one instructor noted that by using warm-ups, graphic organizers, and anticipation sets, she was able to quickly assess how much background knowledge students had and what resources she needed to use to augment student schema. Another instructor shared that utilizing exit slips gave her a fast and effective understanding of who needed re-teaching and who was ready to move on. All educators at our institution indicated

that informal formative assessment benefited their instructional planning and reduced student stress because the assessments were not graded and were only used for instructional purposes (C. Capone & M. Freday, personal communication, March 23, 2019).

These types of assessments are valuable modes of data collection and offer instructors the opportunity to assess students while maintaining an environment of low stress and therefore reduce emotional barriers to success. Informal formative assessments decreased worry and concern, rarely resulted in an inability to perform a task, and engaged students thereby reducing student indifference and increasing overall student performance.

Authentic and Performance Assessment

A fundamental feature of assessment that promotes learning is one that is authentic (Rawlusyk, 2018). According to Baron and Boschee (1995), authentic assessment is a process where students accomplish a real-life task while demonstrating a specific set of behaviors or skills. This type of assessment allows students an opportunity to exhibit their ability through activities such as interviews, discussion, role-playing, presentations, and group work. These types of activities are beneficial to student emotional function because in many cases the students are working individually. They are able to work with peers and interact in a way that does not require all language areas to be mastered. In some instances, students are also able to prepare for their task in advance thereby reducing the pressure associated with improvised performance. Additionally, these types of assessment integrate experiential learning by inviting students to interact with members of their community, which often may result in higher levels of engagement (Kohonen, Jaatinen, Kaikkomen, & Lehtovarra, 2014).

Instructors at our institution chose to increase this type of assessment in the classroom as a way to decrease traditional emotional responses like anxiety, to eliminate cheating, and to increase overall engagement and achievement. Administrators began by requesting all teachers include at least one experiential or authentic style assessment per term plus include a variety of performance-based assessments including debates, presentations, and in class and out of class interviews. This later evolved to teachers integrating this type of assessment regularly for each unit of study. Instructors included a variety of activities to assess speaking and writing in particular. Table 2 illustrates the assessments used.

Table 2: *Authentic, Experiential, and Performance Assessment*

Field Trips to local nature preserves followed by a discussion of what was observed using appropriate vocabulary.
Visiting the local Performing Arts Center and critiquing art or a performance.
Interviewing local business owners after completing a unit on entrepreneurship.
Reading a novel and then conducting a Book Club discussion.
Grammar Fairs taught by students and attended by students of a lower level to practice a variety of concepts.
Choosing a controversial topic and having two classes debate the two views.

This type of assessment was engaging and motivating for our students. It gave them a real-life situation in which to practice their language skills, and they had a greater sense of accomplishment following the activity. Although these types of assessments require creativity and planning, more traditional methods may include written papers, individual projects and presentations, and group projects or presentations (Rawlusyk, 2011). Authentic and performance assessment resulted in increased engagement and achievement at our institution. Students reported feeling more excited about what they were learning and less concerned about whether their language skills were proficient enough to perform. Additionally, students indicated that this type of learning gave them a sense of community and connected them to the university setting (B. Mohammed & M. Ghubayn, personal communication, October 11, 2018).

Metacognition and Feedback

Self-assessment, peer and educator feedback, and one's ability to self-monitor and reflect on personal learning styles are fundamental factors that influence achievement. These methods of assessment encourage students to judge and evaluate their own performance as well as their peers. It encourages a more active role in one's learning and empowers students to be lifelong learners (Carless, 2015).

According to Rawlusyk (2018), written feedback is a powerful tool when used in conjunction with dialogue, exemplars, discussions, presentations, and explicit instruction on how to use feedback to improve achievement. Feedback must be used in a timely and specific way in order to successfully impact student growth and achievement. At our institution, we began using