

**University of Iceland**  
**School of Humanities**  
**Department of English**

# **The Efficacy of Authentic Assessment**

*A Practical Approach to Second Language Testing*

**B.A. Essay**

**Brynja Björk Reynisdóttir**

**Kt.: 051293-3479**

**Supervisor: Ásrún Jóhannsdóttir**

**May 2016**

## Abstract

Second language learning is becoming prevalent in many societies as cross-cultural interaction is increasing, and students must have sufficient language skills to participate in an unfamiliar educational setting. Many pedagogical professionals now recognize the shortcomings of traditional testing methods in regards to children of all age groups, and the need for alternative testing methods. In regards to this thesis, authentic assessment is used to describe different forms of testing that consider student knowledge, success, incentive, and perception of authentic classroom activities. Therefore, this paper explores the efficacy of authentic assessment in the multilingual inclusive classroom. Specifically, this report demonstrates the benefits of authentic testing by discussing definitions of authenticity, validity, task-based assessment and self-assessment, student perception, as well as the implications of incorporating authentic methods. Although scholars define the task types differently in relation to one another, this report aims to give a general definition of authentic assessment and identify practical examples of authentic testing methods that can be used in the language classroom. Furthermore, by discussing the role of the instructor and the importance of instructor training and implementation, this report seeks to define a more holistic approach to language teaching and assessment. In a practical sense, implementation of authentic methods is often problematic. Therefore, this model needs to be given further consideration, especially into the way in which authentic assessment can be implemented into the English as a second language classroom in Iceland and abroad. Not only does this thesis provide evidence to the effectiveness of authentic methods, but it also describes specific methods that are applicable in the second language classroom. An analysis of different authentic assessment methods indicates that alternative testing is useful in the inclusive second language learning classroom.

*Keywords:* authenticity, assessment, holistic, task-based learning, student perception, inclusive learning, second language, instructor training, self-assessment, standards-based learning

## Contents

<b>1. INTRODUCTION.....</b>	<b>4</b>
<b>2. TYPOLOGY .....</b>	<b>6</b>
2.1 TRADITIONAL TESTING METHODS.....	6
2.2 INCLUSIVE SECOND LANGUAGE CLASSROOM .....	8
2.3 AUTHENTIC ASSESSMENT .....	9
<b>3. THE FOUR SKILLS.....</b>	<b>11</b>
3.1 READING.....	12
3.2 SPEAKING .....	13
3.3 LISTENING .....	14
3.4 WRITING .....	15
3.5 CONCLUSION.....	16
<b>4. STUDENTS' PERCEPTION.....</b>	<b>18</b>
4.1 VALIDITY AND RELIABILITY OF STUDENT PERCEPTION.....	19
4.2 STUDENT SELF-ASSESSMENT.....	20
<b>5. INSTRUCTOR TRAINING/IMPLEMENTATION .....</b>	<b>22</b>
5.1 GRADING PARADIGMS .....	22
5.2 IMPLEMENTATION .....	23
<b>6. CONCLUSION .....</b>	<b>27</b>
<i>References.....</i>	<i>29</i>

## **1. Introduction**

Educational and pedagogical theories are constantly shifting. The importance of authentic assessment has become prevalent in recent years, and there is still a vast amount of research to be done in regards to authenticity. The gradual shift from traditional to authentic testing methods demonstrates the attention that should be given to student needs. In this way, instructors may be better able to consider the needs of each student, and tailor a program that targets and utilizes the strengths of each student. Moreover, it is important to consider why instructors have moved away from traditional testing methods, and what methods can be used instead.

In discussing authentic testing methods, this thesis will utilize research and scholarly articles that emphasize the need for methods that incorporate the four skills including; reading, writing, speaking, and listening. According to Bachman (2002) the four skills can be used to assess the student's knowledge of the second language as well as the way in which the students use the second language productively and receptively. This in turn leads to a discussion of the validity of authentic methods as well as the way in which instructors can most effectively incorporate these methods into their teaching.

Additionally, Coelho (2012), a pedagogical specialist with emphasis in multilingual education, addresses ways in which inclusive learning can be utilized by the instructor, and how inclusive learning correlates to four different methods of assessment. Research suggests that it is important to consider inclusive education and holistic learning while considering authentic assessment. Not all students will learn the same way, or be at the same level as their native peers. Cultural differences will affect the way in which students perceive and accomplish their learning outcomes and goals. According to Coelho, an effective inclusive classroom can be used to identify the specific cultural and educational needs of each student.

Training and collaborative support from other instructors and administrators is valuable when considering authentic assessment. In this way, instructors may better be able to embrace student diversity, and use this diversity as a tool in the classroom. Thus, authentic assessment methods can be used to better address the needs of each student. This learning approach aims to welcome all students, while also respecting their unique cultural experience, and use their differences to enlighten and instruct their peers.

As second language and multilingual learning are becoming widely discussed topics in many educational settings, it is important that instructors consider different methods of assessment for both native and non-native speakers. Sambell, McDowell, and Brown (1997) show that students are becoming more perceptive towards authentic methods, and that these methods are viewed favorably over more traditional ones.

In chapter two, the concepts are described in focus; specifically the differences between authentic and traditional testing, a definition of inclusive learning, and the definition of authentic assessment. In chapter three, the four skills are discussed (reading, writing, speaking, listening), and specific methods are outlined to provide ways to incorporate authentic methods into second language learning. In chapter four, student's perception and validity are examined in regards to authentic assessment. Chapter five follows a similar set-up to chapter four, but instead discusses instructor perception and implementation. The sixth chapter addresses implications that come about with authentic assessment, and discusses problems that may arise with second language authentic testing.

Authentic assessment is beneficial in most learning contexts, but especially important in the inclusive language learning setting and authentic assessment is necessary for every age. This analysis looks briefly at standardized testing and the ways in which instructors may incorporate reliable methods to achieve a more holistic approach. However, the primary aim of this report is to provide a practical approach to incorporating authentic assessment into the English second language learning curriculum. Furthermore, this thesis seeks to outline the efficacy of task-based learning approaches and provide specific examples of effective methods of assessment. Consequently, an analysis of different authentic assessment methods indicates that alternative testing is beneficial in the inclusive second language learning classroom.

## 2. Typology

Wiggins (1990) gives a basic definition of authentic assessment by saying, “assessment is authentic when we directly examine student performance on worthy intellectual tasks” (p. 2). On the one hand, traditional assessment is that which relies heavily on inferences about student performance based merely on a small set of standardized test criteria. These tests often require the student to recall specific knowledge and use rote memorization to perform well on tests that only cover a very specific context. Thus, these tests do not assess practical skills or real world ideas, and often do not illustrate student’s abilities to write, listen, speak/discuss, or collaborate, etc. Despite the fact that standardized testing can be used effectively in some instances, it is important that instructors realize that there are many alternatives that provide students with a more holistic learning approach. On the other hand, authentic methods lead to heightened cognitive skills, especially in regards to the task-based assessment methods. Although it is important to address the shortcomings of standardized testing, this report’s aim is not to negate and denigrate standardized testing. It offers alternatives and research that illustrate the importance of authentic assessment, specifically in the inclusive second language learning classroom. Nevertheless, before evaluating studies and recommendations for authentic assessment, one must define the concepts in focus in this thesis.

### 2.1 Traditional Testing Methods

Consideration has been placed in recent years on the shift from traditional methods to authentic assessment and education reform. Traditional testing methods required that students be measured depending on a particular set of standards. These standards test students based on the performance of their peers and a national average. According to Griffin (2007) Glaser published a report in 1963 on criterion-specific testing methods, which was meant to contradict the traditional performance-based methods. Glaser (1963) asserts that students learn across a wide range of interdisciplinary subjects. However, his report emphasizes the need to classify students based on their performance, subsequently testing them merely on pass/fail criteria. Moreover, when Glaser’s report was published, it demonstrated the need to shift from the previous comparative method to one valuing individual achievement rather than the

success of their peers. However, with the pass/fail structure, a very narrow field of criteria was established and the report failed to achieve its preliminary aim.

Glaser (1963) was trying to shift from predictive assessment to a more holistic approach, considering the student on a continuum of little to full fluency. Instead, he created a theory that very closely resembled traditional testing methods. The importance of this historical anecdote is that while researchers and instructors begin to move away from traditional testing methods towards student-centered approaches, it is exceedingly common to develop methods that mirror traditional tests. It is essential that instructors and administrators carefully consider what type of methods they will incorporate into the curriculum, as well as the effect and implications these methods may have on students. Moreover, research suggests that instructors need to be trained to identify the underlying phonological, grammatical, and syntactic systems of each of the task-based assessment types (reading, writing, listening, speaking). In this way, it is easier for educators to develop methods that assess these underlying systems and modify them based on the needs of each student (O'Malley & Valdez Pierce, 1996). Although there is still no strict definition of authenticity and instructors differ in their opinion and use of authentic assessment methods, it is important to note the shift in educational reform in the last few years.

In recent years, educators have been moving from traditional standardized tests to a more authentic assessment based curriculum in many countries. For example, the National Curriculum of Iceland advocates for raising pupils to be well-rounded members of a democratic society. Since 2008, quite a bit of the curriculum advancement has taken place in Icelandic schools. These improvements correlate education with social and cultural aims, provide the opportunity for students to take control of their learning objectives and develop individuality, as well as provide direct criteria for foreign language competency. (Ministry of Education, Science, and Culture, 2012). This cross-curricular viewpoint is similar to the aims of authentic assessment methods outlined later in this analysis. Although there are not substantial studies to assess whether this particular curricular theory is applicable in all contexts, it is a progressive step towards cross-curricular education and authenticity. Furthermore, with the addition of foreign language competency criteria, an emphasis on the needs of foreign language learners and inclusive learning environments is becoming more prevalent.

## 2.2 Inclusive Second Language Classroom

It is of particular importance to consider inclusive education as well as authentic assessment, as not all students will learn the same way, or be at the same level as their peers. According to Coelho (2012), an inclusive classroom is structured to anticipate the needs of all students, regardless of their perceived culture or social customs. Although inclusive learning is a difficult concept to apply, it is one that becomes beneficial in a multilingual classroom. In an inclusive environment, it is important for instructors to embrace student diversity, and employ this diversity in the classroom and their teaching. From a practical standpoint, inclusive learning aims to welcome all students, whether native speakers or newcomers, and use their differences to teach their peers (Coelho, 2012). With a heightened awareness of cultural differences, an instructor can anticipate problems and react to tension that may arise in the classroom. Ambrose et al. (2010) state:

Even though some of us might wish to conceptualize our classrooms as culturally neutral or might choose to ignore the cultural dimensions, students cannot check their socio-cultural identities at the door, nor can they instantly transcend their current level of development... Therefore, it is important that the pedagogical strategies we employ in the classroom reflect an understanding of social identity development so that we can anticipate the tensions that might occur in the classroom and be proactive about them (p. 169-170).

As a result, it is important to note that classrooms are not culturally neutral. Instead, it is important for instructors to consider these different cultural traits and use them as strategies to develop an inclusive learning environment.

Coelho (2012) gives preliminary steps to assess the proficiency of new students in their first (L1) and second (L2) language skills. Although not extensive, these activities provide instructors with a framework for later assessment, and lead into different task-based assessment methods. Included in this model are five assessment steps, the first assesses reading and writing in the student's first language; through labeling, introductory sentence writing, and reading aloud. Once students master the first step, and if they have had some contact with L2, the instructor can then move on to interviewing the student in the L2 language. The interview relies on simple questions to assess basic conversation skills. Third, the student is asked to respond to an oral prompt



in their L2. This activity differs from the interview in the fact that students are asked to produce oral language without the help of questions, but rather respond to one general concept. Fourth, the student is requested to write in L2. Fifth, the student is asked to read in L2 and relate what they read through interviews and prompts. This fifth step aims to use the four previous steps together to form a coherent explanation of what is being read, and correlate them to the student's prior experience and language knowledge. If the student struggles at any point, the instructor then gains knowledge into the areas in which the student needs further instruction. These methods described by Coelho correlate directly to direct authentic assessment, and show that is important for the instructor to consider different methods in the language learning classroom to assess the needs of each student.

Inclusive learning is the way in which instructors can identify and incorporate authentic methods into their curriculum based on the needs of each student. As previously discussed, traditional methods fail to identify the needs of minority students, meaning inclusive learning and traditional methods are incompatible. Acknowledgment of the need for authentic assessment is the first step towards inclusive learning. However, although it is important to discuss the theory behind inclusive and authentic learning, it is equally important to identify specific authentic methods that may be used to develop the curriculum around the needs of each student.

### **2.3 Authentic Assessment**

Authenticity and inclusive learning are multifaceted; both approaches can be employed by instructors to assess the strengths of each student, as well as their overall need. Moreover, there are different forms of authentic assessment, and it is important to consider that each student will not learn in the same way and that some students may need extra preparation, especially if they are new to the classroom. O'Malley and Valdez-Pierce (1996), educational professionals specializing in second language acquisition, consider this a holistic approach to testing, and they assert that this comprehensive approach can be employed in each of the task-based assessment types. Additionally, while holistic assessment includes pre-determined criteria, the difference is that the assessment method requires that the instructor nurture students to realize their full potential, and progress from basic to proficient skills in their second language.

Additionally, O'Malley and Valdez-Pierce (1996) define authentic assessment by saying, "we use the term authentic assessment to describe the multiple forms of assessment that reflect student learning, achievement, motivation, and attitudes on instructionally-relevant classroom activities" (p. 4). Moreover, authentic types of assessment are those that employ and analyze the productive skills (speaking and writing) and receptive skills (reading and listening) as outlined by Baker (2011). According to Baker, these skills correlate with ability. Many bilinguals and multilinguals may use different skills depending on the context, or may be more proficient in one skill over another. However, Baker considers listening a receptive skill, but opinions differ whether or not listening falls strictly into one category, as it is often used in conjunction with the other skills. Alternatively, Davies (1976) describes the receptive and productive skills, and uses them in relation to one another. Davies defines receptive reading as being able to read a variety of different texts at different levels (i.e. academic texts with a large percentage of L1 vocabulary, poetry in L2, news articles, etc.). This type of reading is associated with listening skills and being able to identify the meaning of spoken vocabulary, and he asserts that a higher level of aural knowledge relates to a higher level of reading. The productive skills (i.e. writing and speaking) are those that are used actively in communication in the L2 language, and these skills may range from knowledge of a few phrases to full fluency. He also stresses the fact that most students will not reach the same fluency in their L2 as L1, they do however understand much more of the spoken language than they can speak.

The typology of authentic assessment shows that a shift from traditional to authentic methods has taken place in recent years. The multilingual classroom is moving towards inclusive learning, and also advocating that students take control of their own learning. The inclusive classroom aids students in becoming well-rounded members of society. Inclusive learning and authentic testing aim to employ student diversity and cultural differences to achieve a holistic approach to language studies. For the purpose of this thesis, these four skills will be analyzed in depth with regards to task-based authentic testing methods. These skills are largely dependent upon the classroom setting, interaction between instructors and peers, as well as the language used in the classroom. However, a basic framework and specific examples of different methods of authentic assessment can be identified and utilized in any L2 classroom.

### 3. The Four Skills

Teaching and assessment methods used will differ between instructors and be largely dependent upon then needs of the students. The four skills (reading, writing, listening, speaking) are important in any classroom and the role of the four skills is vast and constantly developing. In relation to the second language learning classroom, it is important that instructors consider these four skills separately and consider the students underlying knowledge of each skill, as well as ways in which student proficiency may be assessed.

Task-based assessment differs from other assessment methods in that task-based assessment tests both the underlying knowledge of the L2 language as well as the ability to use the productive and receptive language skills (Bachman, 2002). This concept directly correlates to testing validity as task-based assessment calls for the need to address the problems of validity as well as the validity of the effectiveness of the testing method being used. Bachman (2002) also stresses the need for instructors to understand the language skills to employ task-based assessment more effectively. Therefore, he asserts that adequate task-based assessment should not only use the task in a real world communicative setting to assess the syntactic and phonological skills of language learners, but also assess how students perform on each given task.

Mislevy, Steinberg, and Almond (2002) cite Brindley (1994) by saying, “Task-Based Language Assessment (TBLA) is ‘the process of evaluating, in relation to a set of explicitly stated criteria, the quality of the communicative performances elicited from learners as part of goal- directed, meaning-focused language use requiring the integration of skills and knowledge” (p. 477). They agree with Bachman (2002) that the instructor must understand the underlying syntactic structure of the language being taught. However, they look more in depth at the specific definition of task-based authentic assessment. They also emphasize the fact that TBLA not only covers the curriculum, but effective TBLA should also consider the societal effects of language learning, as well as goal setting, and communication. More simply, task-based assessment is that which tests the receptive and productive skill (i.e. reading, listening, speaking, writing), as well as considers the individual cultural traits of each student and how they participate in communicative processes with those around them. In turn, an instructor may be better able to realize the student’s weaknesses and build a curriculum

that better addresses the areas where students need more guidance by identifying and isolating each of the tasks and providing ample time across the curriculum to examine them.

### **3.1 Reading**

According to William Grabe (2009) depending on the phonological and syntactic system of the student's L1 and L2, reading will differ, and this should be considered when instructing a student in their L2. However, although the method of instruction may differ, the type of assessment mainly depends on the student's level of proficiency. Of course, not all assessment types will work for every student age, but a general framework for assessment may be developed for a specific age range. More specifically, Grabe (2009) outlines different goals for the assessment of reading. He stresses that reading assessment can be either beneficial or harmful to the learning process, and while this applies to all four task types, he asserts that extra attention must be given to understanding the uses for reading assessment. It is important to note that he directly addresses the fact that to create proper assessment methods, the instructor must first have an awareness of how reading functions.

Reading is a difficult skill to assess. However Valencia (1990) gives a practical example of an effective assessment method. She provides a portfolio approach, and in this way, reading is developed in conjunction with writing skills. Valencia states "it encourages us to use many different ways to evaluate learning, and it has the integrity and validity tha[t] no other type of assessment offers (p. 338). In this approach, students are able to select their best work to include in their portfolio, which also leads the student to self-assess and consider their strengths as well as areas where they need further guidance.

O'Malley and Valdez Pierce (1996) also aim to apply a practical approach to assessing reading. They give four criteria that enhance student ability and learning outcomes. First, the instructor must allow a large amount of the class time to be devoted to reading. This is similar to Nation's (2013) idea that extensive reading is vital to acquiring new vocabulary. He outlines this concept in his four strands method by saying:

It is useful to look at what can make up such a course across the four strands of meaning-focused input, meaning-focused output, language-focused learning and fluency development. About half of the time in a reading course should be spent on meaning- focused input which should largely be done through extensive reading (p. 205)

While it is important to consider meaning-focused output and input, it is also beneficial to consider specific strategies students use to acquire new language.

O'Malley and Valdez Pierce (1996) emphasize the importance of including teaching that targets the specific strategies students use to acquire new vocabulary through extensive reading. This correlates to Nation's (2013) meaning focused output strand, where students are able to apply their reading to class activities or writing prompts. In this way, instructors not only assess the student's skills in reading, but also their skills in the other productive and receptive areas. Third, students should be able to work with their peers, both during reading tasks and the consequent activities. Last, students have the chance to discuss what they have read with their peers and instructor. In this way, the instructor effectively incorporates listening and speaking into the reading assessment, and employs all four task types.

It is important that instructors realize that students will eventually be faced with many different language tasks through higher education and employment, and inclusion of a broad range of literature is helpful to developing a more well-rounded approach to literacy. This is also true of oral language; as students will need to be able to not only use formal classroom language, but communicative language as well.

### **3.2 Speaking**

In order to have proper knowledge of spoken language it is important to note that oral language usage differs in nature from the other forms of language usage. For example, native and fluent speakers tend to use colloquialisms or shortened sentences when speaking. On the one hand, written language is often academic or professional in nature. On the other hand, oral language is often more fluid, requiring that the instructor assesses not only the communicative properties, but also the underlying phonological and syntactic structure of the language being produced. In a formal classroom setting, instructors face the challenge of assessing oral language despite these factors. O'Malley

and Valdez Pierce (1996) state “information is packed less densely in oral language than in written language, with much more use of phrases and simple sentences” (p, 58). In addition, they argue that to gain fluency in the L2 language, speakers must also be able to listen and attain meaning from spoken language. This means that listening is no longer merely a passive skill, but is actively employed in each of the task types.

Although Nations (2013), a vocabulary specialist, does not address authentic assessment specifically, he gives many practical examples that may be applied while using authentic assessment. In regards to speaking, he states that learners acquire and use a smaller vocabulary while speaking and listening than they do from writing. This is important because instructors must incorporate speaking and listening assessment into the curriculum, and they must also assess how these two skills are used in relation to reading and writing. Instructors must also consider how speaking and listening differ in a formal or everyday communicative setting. Furthermore, it is important that instructors realize that listening is not merely passive, and a lot of syntactic and grammatical information is exchanged while the learner is listening.

### **3.3 Listening**

It is important to note that listening is not a passive skill. Listening is difficult to assess qualitatively, but is a key factor in authentic assessment as it is used as a device for the other three task-based assessment types. Listening often requires interaction between speakers; the listener must interpret what is being said and assign meaning to the grammar and vocabulary used. Relating to this concept,

Richards (1983) used research on native language listening processes to suggest that the basic unit of meaning in oral communication is the proposition or idea. The listener’s task is to determine the proposition in an utterance or speech event. The listener does this by using knowledge of syntax and of the real world. (as cited in O’Malley and Valdez Pierce, 1996, p. 52)

In this way, listening can be directly correlated to communicative processes in the real world context. Therefore, the importance of authentic materials for second language learners cannot be understated.

Vogely (1995) applied the task of listening to three different authentic listening tasks over a two-semester second language course at a small university. Interestingly, she used a Metacognitive Awareness Strategy Questionnaire (MASQ) to acquire student perception. Through learner feedback, she found that students believed to develop their listening skills they had to use background knowledge, recognize the type of text provided, anticipate what information will come next, as well as perceive their own confidence and whether or not they believe they comprehend what they are hearing. While each of the student's perception differed in her study, it is important to consider, as it shows that the students are aware of their level of comprehension and what skills they must employ to perform well on the given task. She illustrates the idea that listeners who are engaged and actively participating can identify when they are no longer attentive towards the listening task, and are able to redirect their attention back to comprehension.

As for the authentic performance method, she chose to use a video series in the L2 with three concurrent recall tests. The students were asked to listen and watch the video once, were not allowed to take notes, and at the end had to write down everything they remember from the video. In the second task, the students were asked to turn away from the screen, so that they were only able to listen to the second video. In the third task, they were asked to watch and listen to a video that doubled in length. After assessing the students based on the MASQ and recall tests, she was able to draw a direct correlation between student awareness and performance. She asserts that the students who felt more confident in their listening skills and ability to understand what is being asked of them, performed better on the listening comprehension tasks. Her research also illustrated the fact that task two was the most difficult, and that visual input correlates positively to aural knowledge. Although visual input is not one of the four task types described in this report, it is important to note that each of the receptive and productive skills require different forms of visual, aural, and oral input for maximum effectiveness, and it is important that instructors build a curriculum that supports different input.

### **3.4 Writing**

Whereas it is important to consider listening when assessing the three other task types, it is equally important to assess written language separately. When incorporating authentic written assessment, the role of the student, the purpose of the writing piece, as

well as the role of the instructor and grading system must be taken into consideration. O'Malley and Valdez Pierce (1996) affirm that traditional testing and instruction methods require that writing is taught separately from the other skills. They assert that traditional style dictated that written language need only be taught in the language learning classroom and could merely be assessed in relation to standardized vocabulary, grammar, and spelling tests. However, instructors are now learning that many different methods of written assessment, as well as cross-curriculum learning, are key to producing a well-rounded learner.

In a practical sense, this means that instructors must incorporate different methods of written assessment. On the one hand, this may include a writing prompt that students are asked to respond to. On the other hand, instructors may ask the student to formulate their own prompt and respond. In this way, students are able to develop skills to formulate their own ideas as well as construct a response. This can also be done in a formal or informal setting. For example, the instructor may require that a student formulates an idea that relates to the general theme of the course material or readings. This ensures that the student uses both receptive (reading) and productive (written) skills. In contrast, the student may be asked to respond more informally to a journal prompt or class discussion. This process differs from the more formal method, but is still important in assessing whether or not a student has the skills to formulate an idea and respond accordingly. Another assessment type that may be used is a peer interview. These forms of assessment require the student to formulate questions, listen to their peers, as well as write a report dictating what they learned (O'Malley & Valdez Pierce, 1996). Although these types of writing differ, they are both important because they require the student use not only the productive written skills, but also incorporate reading, listening, and speaking into their task.

### **3.5 Conclusion**

Although instructors must be able to separate and assess each task-based type alone as well as different types of input, they also must remember that the different types can be used in connection with one another, and listening is often a device used in reading, writing, and speaking. While spoken and written language, as well as reading usage, are important to fluency and acquisition of vocabulary, it is equally important to



note that each of these tasks require the learner be able to listen and construct meaning based on the vocabulary and syntax being provided.

It is important to consider these skills in relation to one another and how they affect the way in which students perceive authenticity. Although student perception is not the greatest factor in deciding whether or not a method is authentic, it can be an important tool that instructors use to gauge the efficacy of the assessment method they have chosen. While traditional methods dictate that writing need only be assessed based on standardized vocabulary, grammar, and spelling tests, it is important for instructors to consider alternative methods as not all L2 students will be at the same level. Regardless of the fact that spoken language is often more fluid and less formal than written language, it is equally important that instructors develop authentic methods to assess spoken language and teach students the proper syntax and grammar. Furthermore, this is also true of listening, as it is no longer considered merely a passive skill, but one used in the everyday communicative process of L2 speakers. Additionally, while the efficacy of reading assessment has been studied for many years, it is important for instructors to consider authentic reading methods in the inclusive classroom, as extensive reading is vital to acquisition.

#### 4. Students' Perception

Student perception becomes a key factor in broadening what is being taught in the classroom, as well as the skills attained. The goal of authentic assessment is to develop a program around the needs of each student. One of the most constructive ways of developing a curriculum in regards to the student is by considering students' perception of both authentic and traditional methods. By allowing the student to become an active part of their assessment and create their own learning outcomes, the instructor gains important information about the student's comfort level and willingness to participate in different methods of assessment.

An emphasis needs to be placed on explaining authentic assessment to the students and working around any areas of confusion. Additionally, whereas the need for proper instructor training and implementation is beneficial, it is equally beneficial to consider the student and how they will perceive each of the four task types as well as their underlying knowledge of the second language. Sambell, McDowell, and Brown (1997) studied student perception of authentic assessment of a two year period and across thirteen case studies, that covered a wide variety of task-based assessment types. They used interviews as well as class observation to collect information. They assert that the study was not aimed at asking students whether they preferred authentic assessment to traditional methods, but rather they focused on perception of specific tasks used in the assessment. Through interpretation of the data Sambell, McDowell, and Brown (1997) found that students often had a negative and stereotypical view of traditional assessment methods. Furthermore, Sambell, McDowell, and Brown highlight that many students expressed the opinion that, from their viewpoint, normal assessment methods had a severely detrimental effect on the learning process. They frequently believed that the quality of their learning was polluted or contaminated because they set out, quite consciously, to achieve second-rate or 'poor' learning for the purposes of a particular assessment point.

Additionally, the majority of students viewed alternative assessment as a positive. They saw alternative methods as promoting their learning as well as helping them to develop strategies for learning the material. They also correlated the alternative assessment methods as a developing part of their learning achievement rather than a means to an end. Many students asserted that by using alternative methods, they were

no longer merely memorizing information and learning to perform well on the test, but rather trying to understand the content presented in their reading and lectures. The fact that the majority of students tested had favorable results when being tested using authentic methods, and felt that their achievement and learning was heightened, demonstrates the efficacy of authentic testing methods. While it is important to consider authentic testing based on both student and instructor perception, it is equally important to discuss the more practical side of implementation. While student perception plays a key role in self-assessment, it can also be beneficial while considering reading, writing, speaking and listening.

#### **4.1 Validity and Reliability of Student Perception**

A correlation can be drawn between the benefits and validity of authentic assessment and how students perceive the assessment methods. While it is important to consider different forms of assessment and how they can be incorporated into the multilingual classroom, it is equally important to address how students perceive each task. In their article Sambell, McDowell, and Brown (1997) consider validity and reliability to be decisive components when considering authentic assessment. Conversely, they also state that there are many other components that make authentic assessment methods important. They further go on to say, “the concept of validity has been extended in recent years to include aspects such as the effects of assessment or testing on the teaching and learning context and the social consequences of the use of assessment information.” (p. 349). Moreover, they use this concept of validity to highlight the importance of acquiring information on students’ perception of different authentic methods. They assert that most scholars and articles that discuss authentic assessment, also discuss validity. They state:

At a general level the definition of validity is unproblematic: at root it is universally used to convey the primary meaning of ‘measuring what you set out to measure.’ This definition is deceptively simple. As Hammersley (1987) recognizes, when we look in more detail at discussions of validity, we do not find a clear set of definitions but a confusing diversity of ideas (p. 349).

Thus, the question that instructors must address then becomes whether or not validity is the sole factor in deciding the need for authentic assessment.

Moreover, Sambell, McDowell, and Brown (1997) study whether certain forms of assessment have an impact on teaching as well as student achievement. Frederiksen (1989) discusses this problem by saying:

Many large-scale testing programmes lead to a damaging narrowing in what was taught due to a concentration on meeting test requirements...an important task for educators and psychologists is to develop instruments that will better reflect the whole domain of educational goals and to find ways to use them in improving the educational process (p. 201).

Many traditional testing methods fail to assess the overall achievement of students, and rather lead to a negative form of assessment in which instructors merely teach students to perform well on the test. As a result, students often do not learn key skills such as problem solving, goal setting, self-guidance, self-awareness etc.

#### **4.2 Student Self-assessment**

While traditional methods fail to assess the overall achievement of students, self-assessment can lead students to a more positive view of authentic methods. While self-assessment may not be beneficial for younger children, self-guided learning becomes important for student in secondary and upper secondary school.

Chen (2008) compares the difference between student perception through self-assessment and the perception of the instructor with regards to the students' performance in oral English language. The article demonstrates that within two sets of self-assessment and teacher assessment, the students learned to assess their learning accurately and outline their goals. The article also demonstrated that the students were asking many of the same questions and had similar feedback to the instructor. This article demonstrates the fact that students are aware of their goals and the learning outcomes they wish to achieve. Furthermore, the fact that there was agreement between the students and instructor shows that self-assessment can be used as an effective tool to aid instructors in learning about the student's strengths and weaknesses.

Most students view authenticity as a positive factor that leads them to better self-assessment and helps develop skills in goal setting. They feel that this learning style helps them to develop a holistic approach to learning, rather than one that assesses very

specific criterion. Thus, the discussion of validity and reliability concerning student perception becomes valuable in determining what methods students perceive as positive additions to their learning. In addition, the discussion of validity leads instructors to consider the importance of student self-assessment and awareness of their own learning outcomes. Conversely, when considering authenticity, instructors need training to identify effective methods of implementation and experience in how to combat traditional grading paradigms.

## 5. Instructor Training/Implementation

Until recently, educational theory has been largely dominated by traditional learning and assessment methods. This may be because many instructors do not have the skills necessary to implement an effective authentic curriculum. The importance of identifying and combating traditional paradigms cannot be underestimated. The design of the authentic method used will differ greatly depending on the instructor and the students. Thus, it is important to consider different methods of instructor training and how instructors may more effectively implement authentic methods.

### 5.1 Grading Paradigms

For many years traditional grading paradigms have been accepted and undisputed across the curriculum. Vatterot (2015) outlines these common paradigms and offers alternatives to these traditional methods. She states:

Equating remembering with learning and with being smart is a long-standing tradition...Given the pervasiveness of this view, it is not surprising that many teachers still fixate on the coverage of rote knowledge and still equate remembering with learning. Often requiring students only to know and understand, many teachers have not expected students to function at the levels of application, analysis, and synthesis. Many teachers are not sure how to teach for higher-level thinking, and many students do not know how to think at higher levels because they have not been expected to do so (p. 27)

Vatterot's standards-based paradigm is very similar to the task-based approach. The goal of both methods is not to assess students based on what they know, but more specifically how they apply what they know in an everyday context. Recently there has been a shift from the traditional paradigm of merely knowing and understanding important facts to a more holistic approach. This means students need to be able to apply, reflect, and evaluate new knowledge. This shift brings about the need for authentic methods of assessment. The task-based assessment method becomes important in the fact that it outlines and defines specific needs related to the second language learner, as well as identifies different approaches instructors may take in addressing foreign and second language competency. It is also important for instructors to

recognize traditional paradigms, and learn how to combat them through alternative teaching methods and authentic assessment methods.

Simply put, these traditional paradigms are still pernicious in the educational system today. While a shift from tradition to higher levels of thought has taken place in recent years, many schools still struggle to break free from traditional paradigms. In contrast, traditional paradigms include ideas such as students being judged with extrinsic grades while still learning, creating a fear of failure and an environment where students merely reiterate information that is agreeable to the instructor, rather than developing their own opinions. Also, failure is the ultimate judgment of a student's ability, and deciding factor in higher education. Contrastingly, according to the standards-based grading model, mistakes and failure is a natural part of the learning process, and lack of understanding is not an indication of inanity, but rather an indication as to what areas students require more support.

Thus, an approach that focuses solely on teaching for the test is not as effective as authentic methods of assessment. It is important that instructors develop their curriculum around the pupils and construct the student as the center of the classroom. It is equally important to consider the fact that long-term knowledge is often not gained from memorization alone, but rather from hands-on experience and practical application. In this case, proper instructor training is key to proper implementation of authentic methods.

## **5.2 Implementation**

Another factor that must be considered is the implementation of authentic assessment. It is important to consider the way in which instructors design authentic assessment to include in their classroom. In theory, authentic assessment is important in the language learning classroom. However, in practice, authentic assessment can often become problematic. To address these problems specifically Gulikers, Bastiaens, and Kirschner (2004) provide an outline for authentic assessment. They claim that in previous years, education has been moving from achievement based tests that merely emphasize students' ability to memorize, to a system that is now placing an emphasis on training qualified students with a wealth of knowledge in different areas. In this way, it is important for instructors to consider the student's education across the curriculum.

This is especially important when assessing second language users, as they will perform differently than native speakers, and often will require more attention in the classroom.

In this case Gulikers, Bastiaens, and Kirschner give a step by step approach to implementation that identifies five necessary objectives the instructor must consider when incorporating authentic assessment. The first is the assessment task, whether it is reading, writing, listening or speaking, it is crucial that the instructor constructs activities that are meaningful to the student and that test their knowledge of the subject matter. Second is physical context. Simply put the context in which the learner uses their knowledge of the L2. The communicative process is dependent upon whether or not the learner is using formal or colloquial language, as well as whom they are addressing. Third is the social context. Whether or not a student is able to work comfortably with their peers and instructor as well as address those around them in their L2 through different activities. Fourth is the assessment results; this is also dependent upon the way in which the instructor interprets the results and directly uses them to address student weaknesses in the curriculum. Moreover, the instructor must consider whether or not the results were taken from a wide range of different testing methods and compared with one another to acquire a holistic result. Fifth is the assessment criteria and standards. The authors cite Arter and Spandel (1992) by saying “criteria are those characteristics of the assessment result that are valued; standards are the level of performance expected from various grades and ages of students” (p. 75). More specifically, an authentic criterion is that which the students can understand fully in regards to their own assessment (i.e. they are fully aware of how they are being assessed and what is expected of them), as well as be able to understand them in the physical and social contexts.

It is crucial that instructors consider these five objectives when considering how to implement different assessment methods in their curriculum, as well as how these contexts affect the way in which students will use their knowledge later in life. According to O'Malley and Valdez Pierce (1996) it is equally important to consider teacher training when considering implementation. The important factor for instructors to consider is that because there are so many different forms of authentic assessment, they must consider the ones that are most applicable to their curriculum and for their students. Many of these authentic approaches can be specifically adapted depending on



the needs of the instructor and student, and it is equally important that the administration considers how to train instructors to identify areas of need as well as how to build their lesson plans around authenticity. However, it is equally important for the school districts to reach a consensus on what types of assessment may be useful, as instructors with differing levels of experience and interest in authenticity will have differing viewpoints. Finally, O'Malley and Valdez Pierce advise instructors to work collaboratively with their colleagues to enhance their knowledge and experience, especially while trying new modes of assessment. Instructors and administrators must decide how they will implement these new testing methods into the language learning classroom. However, they must also consider the validity of these methods, and whether or not they can be effectively used when teaching in a second language.

### **5.3 Instructor Validity**

Validity is one important factor when deciding whether an assessment method is authentic or not. Wiggins (1990) states, "test validity should depend in part upon whether the test simulates real-world 'tests' of ability. Validity on most multiple-choice tests is determined merely by matching items" (p. 2-3). More simply, for an assessment method to be valid, it must effectively evaluate what it was designed to evaluate. In the simplest terms, a reading test is generated to assess reading and a writing test to assess writing. However, this definition is often problematic, as the instructor must consider many different factors while addressing student performance.

Gipps (1994) addressed performance with regards to the four types of validity. The first is predictive validity, how the testing method predicts whether a student will be successful in higher levels of education. The second is concurrent validity, or in simpler terms, whether or not the testing method can be compared and used in concurrence with other tests that assess similar skills. The third is construct validity, or whether the test interprets and evaluates the basic structure (whether syntactic, grammatical, phonological etc.) of the skill being assessed. Fourth is content validity, or does the curriculum cover the material needed to do well on the test. However, these four types are problematic, because they fail to address the different cognitive skills used, as well as the fact that they fail to recognize the fact that L2 users will not learn at

the same pace or level as their native peers, and this brings about implications for authentic assessment.

In fact, the curriculum often fails to address the needs of the L2 student. This is why it is important to note the similarities between the aims set by the curriculum and the objectives being tested, and whether or not they account for the individuality of each student. It is important to consider the validity of the curriculum learning outcomes and how they correlate to the testing content. Furthermore, instructors must consider whether these learning outcomes mirror what is being tested, and whether or not the testing methods accurately assess the areas in which students falter and may need additional guidance. Specifically, it is important to assess how students acquire vocabulary in a second language through different tasks (i.e. reading, writing, listening etc.) and what methods aid students in gaining fluency in their second language. Furthermore, authentic testing aids instructors in determining student knowledge, and the classroom focus shifts to activities that support and consider individual student needs. The last important factor to consider in regards to validity is how the assessment method affects the instructors and students, and consequently how the assessment can be beneficial in the classroom. (O'Malley & Valdez Pierce, 1996). Discussions of validity, in turn, lead to consideration of the different task-based assessment types, and the implications that may arise from these testing methods.

Research suggests that instructors need to be able to consider their own implementation as well as how validity and traditional methods affect student learning. All of these factors previously discussed affect the way in which instructors and students perceive authentic methods. Additionally, whereas the efficacy of authentic assessment is apparent, there are many implications that arise with the implementation of authentic methods.

## 6. Conclusion

Whereas authentic assessment is beneficial in the language learning classroom, it is also important for instructors to consider the implications of implementing and authentic framework. Specifically, it is important for instructors to realize that a great deal of planning and training must be undertaken when implementing authentic assessment into the classroom. It is also important that instructors consider the context in which each of the task-based assessment types are used, in this way he/she will better be able to form assessments that benefit all of the students. For example in regards to oral language, O'Malley and Valdez Pierce (1996) stress that it is beneficial that instructors consider language as it is spoken in an everyday conversational manner, rather than it is usually represented in a prescriptive and formal written context. This process is two-fold; first the instructor must recognize the difference between written and oral language, and second be able to differentiate and create assessment methods that address each type individually. In theory, this concept is easily explainable, but in a practical context, it is difficult to address the needs of each student in regards to oral and written language. The difficulty arises from the fact that often written and oral language are intertwined, and not easily separated from one another.

It is difficult to construct a practical and authentic approach to assessment in any classroom, but becomes especially difficult in the L2 classroom. Instructors must consider the validity of the assessment method when considering whether or not to incorporate it into the classroom. As stated in the validity section of this analysis, the most prevalent form of validity is explained by the four types (predictive, content, construct and concurrent). However, each of these validity types are problematic, and lead many instructors to teach content that only relates to very specific tests.

In the case of predictive validity (prediction of success on performance tests) most often only those that pass the tests go on to higher levels of learning, and it is difficult for instructors to assess the areas of weakness as well as the strengths of the students who did poorly. Therefore, instructors are not able to develop strategies to guide the students who perform poorly, and they, in turn, continue to perform poorly on future tests. While these tests inform the student that they either did well, average, or poorly they do not show the student why they did or did not do well. The aim of authentic assessment is to show the student and instructor areas where more help is

needed, as well as get the students into the habit of controlling their own studies. This is also the case with construct and concurrent validity. As for concurrent validity, if one testing method is flawed all of the testing methods being compared will ultimately become flawed. Often traditional tests do not assess the underlying construct of the skills that need be acquired to learn a new language (Gipps, 1994). In this way, task-based authentic assessment becomes important, because not only does it test student performance in the four areas, it also prompts instructors to consider the underlying structure of the language and how each skill works together.

Alternatively, while instructors can use each of the four skills in relation to one another, it is often difficult to separate them when needed. Reading and writing often work in conjunction, as do speaking and listening. This too can be problematic, as instructors must take into consideration what effect the student's native language will play on the learning process, as well as consider different methods that may be employed to test the student's skills accurately in the second language. (O'Malley & Valdez-Pierce, 1996) As discussed above, William Grabe (2009) stresses the fact that although reading assessment (and any task-based assessment) can have many different objectives, it is important to be mindful of the effect that the specific assessment method has on the skill being tested.

This thesis seeks to outline and correlate the efficacy of task-based learning approaches and practical methods of assessment in the second language learning classroom. It further shows that different authentic assessment methods are beneficial in the second language classroom. This report is a brief look at the efficacy of authentic methods both in the Icelandic classroom and globally. However, it is important to give further consideration to authentic assessment, especially into the way in which these methods can be applied in the English as a second language classroom in Iceland and abroad.

## References

- Ambrose, S. A., Bridges, M.W., DiPietro, M. & Lovett, M.C. (2010). *How learning works: Seven research-based principles for smart teaching*. San Francisco, CA: Jossey Bass.
- Arter, J. A., & Spandel, V. (1992). An NCME instructional module on: Using portfolio of student work in instruction and assessment. *Educational Measurement: Issues and Practice*, 11(1), 36-45
- Bachman, L. (2002). Some Reflections of Task-based Language Performance. *Language Testing*, 19(4), 453-476.  
doi: 10.1191/0265532202lt240oa
- Baker, C. (2011). *Foundations of bilingual education and bilingualism* (5th ed., p. 1-10). New York, NY: Multilingual Matters.
- Chen, Y. (2008). Learning to Self-Assess Oral Performance in English: A Longitudinal Case Study. *Language Teaching Research*, 12(2), 235-262.  
doi: 10.1177/1362168807086293
- Coelho, E. (2012). *Language and learning in multilingual classrooms* (p. 1-160). New York, NY: Multilingual Matters.
- Davies, N. (1976). Receptive Versus Productive Skills in Foreign Language Learning. *Modern Language Journal*, 60(8), 440-444. doi: 10.1111/j.1540-4781.1976.tb03667.x
- Frederiksen, J. R., & Collins, A. (1989). A Systems Approach to Educational Testing. *Educational Researcher*, 18(9), 27-32. doi: 10.3102/0013189X018009027
- Gipps, C. (1994). *Beyond testing: Towards a theory of educational assessment*. (p. 58-75). London. Falmer Press
- Glaser, R. (1963). Instructional Technology and the Measurement of Learning Outcomes: Some Questions. *American Psychologist*, 18(1), 519-521.

- Grabe, W. (2009). *Reading in a second language: Moving from theory to practice* (p. 108-390). New York, NY: Cambridge University Press.
- Griffens, P. (2007) The Comfort of Competence and the Uncertainty of Assessment. *Studies in Educational Evaluation*, 33(1), 87-99.
- Gulikers, J. T., Bastiaens, T. J., & Kirschner, P. A. (2004). A five-dimensional framework for authentic assessment. *Educational technology research and development*, 52(3), 67-86.
- Hammersley, M. (1987). Some Notes on the Terms 'Validity' and 'Reliability'. *British Educational Research Journal*, 13(1), 73-81.
- Ministry of Education, Science and Culture. (2012). The Icelandic national curriculum guide for upper secondary school : general section (2012). In *National Curriculum Guides*.
- Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2002). Design and analysis in task-based language assessment. *Language testing*, 19(4), 477-496.  
doi: 10.1191/0265532202lt241oa
- Nation, P. (2013). *Learning vocabulary in another language* (2nd ed., p. 180-205). New York, NY: Cambridge University Press.
- O'Malley, M., & Valdez-Pierce, L. (1996). *Authentic assessment for English language learners* (p. 1-269). Addison-Wesley Publishing Company.
- Rae, A., & Cochrane, D. (2008). Listening to Students: How to make written assessment feedback useful. *Active Learning in Higher Education*, 9(3), 217-230. doi: 10.1177/1469787408095847
- Richards, J.C. 1983. Listening and Comprehension: Approach, Design, Procedure. *TESOL Quarterly* 17(2): 219-240. doi: 10.2307/3586651
- Sambell, K., McDowell, L., & Brown, S. (1997). But is it Fair? An Exploratory Study of Student Perceptions of the Consequential Validity of Assessment. *Studies in Educational Evaluation*, 23(4), 349-371.

- Valencia, S. (1990). Assessment: A Portfolio Approach to Classroom Reading Assessment: They Whys, Whats, and Hows. *The Reading Teacher*, 43(4), 338-340.
- Vatterott, C. (2015). Rethinking grading: Meaning assessment for standards-based learning. (p. 26-40). Alexandria, VA: ASCD
- Vogely, A. (1995). Perceived Strategy Use During Performance on Three Authentic Listening Comprehension Tasks. *Modern Language Journal*. 79(1), doi: 10.1111/j.1540-4781.1995.tb05414.x
- Wiggins, G. (1990). The case for authentic assessment. *ERIC Digest*. 12(1), 1-6