Ben’s Story
Authentic Task-Based Achievement Test

Karina Lopez
# Contents

Contents ................................................................................................................. 2  
Introduction ........................................................................................................... 4  
Project Description ............................................................................................... 4  
  Background Information ..................................................................................... 4  
  Host class .......................................................................................................... 4  
  Host Institution .................................................................................................. 4  
  Group members ................................................................................................... 5  
Language Assessment Instrument .......................................................................... 6  
  Type of Assessment .............................................................................................. 6  
  Purpose ............................................................................................................. 6  
  Item Design Approach .......................................................................................... 6  
  Test Type .......................................................................................................... 7  
  Scoring Approach ............................................................................................... 7  
Specification Information ....................................................................................... 8  
  Specifications ..................................................................................................... 10  
  Theme ............................................................................................................... 10  
  Objectives ......................................................................................................... 10  
  Specification ....................................................................................................... 10  
Results ..................................................................................................................... 15  
  Volunteer Trial Test Taker Results ...................................................................... 15  
  Trial of NNS ....................................................................................................... 15  
  Trial of NS ......................................................................................................... 15  
  NNS vs. NS Test Trials Reflection & Future Considerations ............................... 16  
Target Test Taker Results ....................................................................................... 18  
Score Analyses ...................................................................................................... 18  
  Item Analyses .................................................................................................. 19  
Reflection and Discussion ..................................................................................... 20  
  Language Assessment Concepts ....................................................................... 20  
  Components of the Test ..................................................................................... 22  
  Validity ............................................................................................................. 22  
  Context Issues .................................................................................................. 23  
  Issues of Test Development .............................................................................. 24  
Future Inquiries .................................................................................................... 26  
References ........................................................................................................... 27  
Appendices ........................................................................................................... 28
Assessment Project Paper

The project paper discusses relevant concepts of language assessment in the context of the assessment project entitled Ben’s Story and a literature review on Authentic-Task & Performance-Based Assessment.
Introduction

The following paper provides information about the target audience, location, and participating assessment project designers. The paper also includes the specifications of the achievement test, the test itself, the student results, and a reflection and discussion. The reflection focuses on the strengths and weaknesses of the Assessment project. The paper ends with questions for future inquiries.

Project Description

Background Information

Host class

The students at McKinley Community School for Adults (MCSA) are from different Asian backgrounds. There are Chinese, Japanese, Korean, Taiwanese and Thai learners. The students are between twenty and eighty years old. They have all taken a proficiency placement test placed at level 4. ESL 4 is commonly known as mid-intermediate level learners. Several of the students have studied at MCSA before, and others are new. The learner’s goals range from aiming to effectively communicate in society, to attending an American university. There are a total of thirty two students in the class.

Host Institution

McKinley Community School for Adults (MCSA) in Honolulu, Hawaii. Classes run for a full three month semester. The class meets Monday through Thursday, from 8:00 to 11:00 am. The classroom number is 104 at all times. The school requires every to attend Lab time, where students work with a program called Achieve 300 (formerly known as Empower 3000). Lab time consists of forty five minutes, of reading an article, answering reading comprehension questions, and responding to a writing prompt online.
All students are required to purchase a grammar book and a text book. The book series used for this course is called *Stand Out, Standard Based English* by Staci Jenkins & Rob Jenkins (2008). The book belongs to a series that includes all five ESL levels.

**Group members**

The team members for the Assessment project are Karina Lopez, Madoka Ikeura, and Martin Molden. In this project I was the host teacher who has spent three months with the students, working at McKinley Community School for Adults (MCSA). I am a student at Hawaii Pacific University, in the Masters of Arts program majoring in Teaching English as a Second Language. I have also taught Japanese to young true and false beginners of English. In the past I have taught Spanish, Art, and Swimming. Martin Molden, also a participant of the Masters of Arts program majoring in Teaching English as a Second Language, has taught primary education classes in Norway. Martin Molden has also made plans to teach English as Second Language in New York. Madoka Ikeura has taught English Language in Japan, and hopes to return to Japan to utilize her Masters of Arts in Teaching English as a Second Language.
Language Assessment Instrument

The Authentic Task-Based Achievement Test (ATBAT) entitled Ben’s Story was administered on March 23rd, 2015. The test is based on authentic tasks and utilizes gap-fill, multiple choice, short answer questions, and extended answer in response to a writing prompt. Students are asked to read a narrative based on a fictitious character name Ben and his shopping experience. Students are then asked to read a dialogue, follow written instructions of a dialogue, and interpret a map. Part four of the test provides a conclusion of the characters’ Bens’ experience and students are asked to write a complaint letter pretending to be Ben (view appendix A)

Type of Assessment

Purpose

Every Thursday of the week at MCSA, at 10:00 am, class ESL level 4 is administered an achievement test to test the knowledge Ss gained during the week. The test exists in a format and theme language learners have been using all week and is familiar to them. In the context of the achievement test Ben’s Story, language learners were working on a unit entitled Community. In this unit Ss covered locating local services by asking friends for advice, interpreting a google map, following and giving directions, completing a refund transaction with a clerk at a store, and writing a complaint letter. The purpose of Ben’s Story, was to gather information on how much language learners have learned and to test tasks like interpreting a map and writing a complaint letter.

Item Design Approach

The approach that was taken in designing each item for each part was the “simple” approach. It is called the simple approach, because items could not be designed to be complicated, quiz-like, or mysterious for the target language students. The items were designed to be simple, short, and consistent, because the target language students are very sensitive to tests, complex items, and long tasks.
Test Type
The test consists of four very different yet consistent and connected parts. Part one consists of reading a theme based story and answering communicative and comprehension questions. In part two, students (Ss) have to read and comprehend a dialogue, follow written directions, and interpret a google map. The third part, consists of a multiple choice activity, where Ss must complete sentences choosing words from a word bank. Finally, in part four Ss are given a scenario and must write a complaint letter. All four parts are connected by the theme and narrative.

All four parts of the test are integrative yet direct. Huges (2003) says, “Direct testing implies the testing of performance skills, with tasks as authentic as possible, tests that test directly, test the skills that we are interested in fostering, then practice for the test represents practice in those skills” (p. 54). In parts 2 and 4 language learners are performing an authentic task that was taught to them in class.

The integrative test type is a test that utilizes more than one linguistic elements in the accomplishment of a given task. In the context of the test (Ben’s Story), language learners are required to utilize multiple language skills and focus on a specific language tasks. For example, in part 3 of the test (Ben’s Story), Ss are required to read a story with missing words, recognize the correct word to use, and actually know how to use it, to correctly complete the vocabulary task. Furthermore, the test is also integrative, because it utilizes reading comprehension in all four parts, yet still tests for specific language tasks.

Scoring Approach
According to Hughes (2003), “if a test is to have validity, not only the items but also the way in which the responses are scored must be valid” (p.32). The scoring approach for Ben’s Story is criterion based, and utilizes two individual detailed rubrics to score all each part. The other two parts of the test have scoring keys. The scoring style is analytical, because of the rubrics and answer keys designed to score the sections. The test is scored in sections, and then the results of the section are added together to give the final score. The score exists in number
format, for example 80% out of 100%. Whereas if the test were scored holistically, there would be no numerical scores, but a scale of that utilizes whether the Ss do excellent, okay, or poor. The scoring is valid, because the rubrics utilized for each test section aims to score distinctive language tasks. The rubrics and scoring keys support the validity of the test, because each item sticks to one specific type of answer. For example part 4, Ss have to write a complaint letter. The rubric for part 4 scores how well Ss were able to express a distinctive opinion, idea, or part of the story and fulfill a requirement of the complaint letter format. The rubric used for all four parts of the test are task specific, and do not test for grammar, because the test is not made for testing grammar.

**Specification Information**

**What testing approach should be suitable to the student population (needs, goals, interests, comfort level with 'tests')?**

Students (Ss) are between the ages of 20 and 82 years old. The students are of different backgrounds that include Korean, Chinese, Japanese, and Thai. Some of the students are also retired and interested in further improving their English language acquisition to communicate with family and friends. Other students are interested in improving their English language skills as a means to function in society. The younger students in the class seek to improve their English language skills to be able to obtain employment and future admission into an American university. As a result, the format and technique of the test needs to be familiar, theme based, authentic, and contextualized.

**To what extent can you make your test fun and attractive to the learners (content, task, use of authentic materials, etc.)?**

Tests must be short and clear for the target Ss at MCSA. Students respond negatively to long and tedious tasks, or to extensive multiple-choice questions that test complex
ideas, heavy content, and multiple grammar features. On the contrary, students respond positively to colorful images, short questions and relevant tasks.

The following techniques and tasks chosen for the test are short, include images, and utilize relevant content. The theme of the test is shopping related, and has four sections. Each of these sections caters to a different objective and a different task. Furthermore, the test assesses reading comprehension (part one), written directions and map interpretation (part two), word recognition skills (part three), and writing skills that include producing and organizing of information from the text, producing an organizing of complete and relevant ideas, opinions, and positive advice. It is also a goal to make the tasks integrative and communicative in order make the whole test relevant, purposeful, and meaningful.

Continue onto next page...
Specifications

Theme: Shopping

Scenario description: The character Ben is shopping for bread at FoodCountry in Waipahu, Hawaii (Part 1). Ben and his wife are having a conversation in which he relays his bad shopping experience, Ben’s wife then provides instructions on how to get to another FoodCountry (Part 2). Ben arrives at FoodCountry and makes a return (Part 3). As a solution, Ben decides to write a complaint letter to FoodCountry about the negative shopping experience at FoodCountry (Part 4).

Objectives
Students will be able to:
- demonstrate reading comprehension skills
- follow written direction and interpret a map
- demonstrate vocabulary recognition knowledge
- communicate a complaint about a retail transaction and provide positive advice

Specification
1. Content

   a. Operations (tasks for learners):

      Section 1: Reading comprehension. Skim or scan texts for specific ideas. Guess meaning of unknown words from context. Understand statements. Respond to questions with information from the text and personal ideas.

      Section 2: Locating local service on a map, based on a written dialog and instructions.

      Section 3: Vocabulary. Recognizing missing vocabulary items in a passage containing simple sentences.

      Section 4: Writing a complaint letter. Expressing dislikes and discomfort about a recent event, and providing positive advice (section 1) through a letter.

Continue onto next page...
b. Types of text:
   Section1: Narrative
   Section2: Dialogue between two people
   Section3: A narrative containing gapped simple sentences with accompanying word bank
   Section4: narrative scenario and complaint letter

  c. Addressees of text: Adult non-native speakers of English, from all different Asian backgrounds

  d. Length of text:
   Section1: One to two paragraph long
   Section2: 10 to 15 simple sentences
   Section3: 10 sentence long
   Section4: 5 to 7 sentences

  e. Topics: Everyday
   Section1: Buying bread and discovering a problem
   Section2: Locating local services in a map
   Section3: Making a return
   Section4: Complaint about bad service

  f. Readability: Fresh reading ease: 85.0, Flesch-Kincaid Grade level: 3

  g. Structural range:
   Section1-4: Simple sentences

  h. Vocabulary range: beginning level, everyday

  i. Dialect, accent, style:
   Section1-3: North American English, colloquial
   Section4: North American English, formal

  j. Speed of processing:
   Section 1-3: 50 to 60 words per minute (reading speed).
   Section 4: 20 words per minute (writing).

  2. Structure, timing, medium/channel and techniques

    a. Test structure: 4 sections
       Section1: Reading comprehension
       Section2: Reading comprehension of a dialogue and map interpretation
       Section3: Vocabulary recognition
       Section4: Writing a complaint letter
b. Number of items:
   Section1: 5 items
   Section2: 3 items
   Section3: 10 items
   Section4: 1 item

c. Number of passages:
   Section1-4: No passages, narrative text constructed for the purpose of this test

d. Medium: Pencil-and-paper

e. Timing: 10 minutes per section = 40 minutes + 5 minutes for reading instructions = 45 minutes total

f. Techniques:
   Section1: Extended response
   Section2: Gap filling
   Section3: Multiple-choice in gap filling format.
   Section4: Extended response

3. Criterial levels of performance

   a. Criteria:
      Students receiving 80-100 points receive mastery (©)
      Students receiving 79-60 points receive acceptable (√ check)
      Students receiving 59-0 points receive review (R)

   b. Scoring procedure:
      Each of the four sections is rewarded 25 points. Total points are 100.
      Number of Raters: 1
Part 1

Rubric

- (5) Complete sentence with relevant information
- (2.5) Incomplete sentence with relevant information
- (0) Complete or incomplete sentence with irrelevant information

Answer Key

1. Student proposes that something capable of leaving bite-marks has taken a bite out of the bread.
2. Student proposes that Ben needs the receipt to return the bread.
3. Students explain why or why they wouldn’t eat Ben’s bread backing it up with content in the reading.
4. Student envisions that Ben does something that would be likely given the context.
5. Student answered the question (yes or no) and provided a description if applicable.

Part 2

Answer key

- (5) The student placed the X correctly
- (5) The student has drawn a route that leads to the X
- (5) The student’s route goes through Nuuanu Avenue towards Foster Botanical Garden
- (5) The student’s route goes through North School Street
- (5) The student’s route ends after following North School Street for two block

Part 3

Answer key

a. (2.5) Clerk  j. (2.5) store credit
b. (2.5) Issue  i. (2.5) Cash
c. (2.5) Return
d. (2.5) Refund
e. (2.5) Items
f. (2.5) Purchased
g. (2.5) Wallet
h. (2.5) Receipt
Part 4

Rubric

<table>
<thead>
<tr>
<th>Content</th>
<th>Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Has the issue been presented successfully?</td>
<td>/15</td>
</tr>
<tr>
<td>- Has a request been proposed successfully?</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Format</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>- Are the address and date located in the top left corner of the letter?</td>
<td>/10</td>
</tr>
<tr>
<td>- Is the receiver’s address specified in the second paragraph?</td>
<td></td>
</tr>
<tr>
<td>- Is there a greeting/salutation leading into the main body of the letter?</td>
<td></td>
</tr>
<tr>
<td>- Are the issues addressed in the main body of the letter followed by a request to these issues?</td>
<td></td>
</tr>
<tr>
<td>- Does the bottom part of the letter contain the following elements in the following order vertically? Closing -&gt; printed -&gt; signature</td>
<td></td>
</tr>
</tbody>
</table>

Total score /25
Results

Volunteer Trial Test Taker Results

Trial of NNS

The following information is about two specific trial volunteers that stood out the most among the nine individuals that were trialed.

Non-Native Speaker Trial #1: NNS-A:

The first non-native speaker the test was trialed on will be named NNS-A. This learner is special, because she is unlike any other trial learner-chosen to trial Ben’s Story. First, NNS-A is very close to the age of the target test takers at MCSA. NNS-A is an older immigrant, and a native Spanish speaker, who learned English later in life. Secondly, NNS-A has taken a placement test with a similar institution like MCSA and placed at middle intermediate level, which at MCSA mid-intermediate levels are 3 and 4.

NNS-A was an excellent candidate to trial the test on, because NNS-A is at a similar production and receptive English skill level to that of the target test takers at MCSA according to the test results and comparison.

Lastly, NNS-A scored a 94% out of 100% on the test Ben’s Story (appendix e). It was later discovered (after the item analyses) that NNS-A made the same error that 16 out of 21 Ss committed taking the test. It turns out, item b was extremely difficult (view appendix b for item analyses, or page 19).

Trial of NS

Native Speaker Trial #2: NS-B:

NS-B will be the name of the second volunteer, who is also older, knows two languages, and grew up in the United States. NS-B is closer to a native speaker of English due to the experience of having lived in the United States so long, and using English the majority of her life. NS-B has a higher education, and scored a
100% on the test. Although the high score is positive, NS-B cannot compare at all to the Target Test Takers (TTT) at MCSA, because she has a longer experience with English, and has been educated solely in English. Consequently, the results of NS-B’s test does not contribute a lot to the developing of Ben’s Story. Rather, as an experienced NS it was shared that the test was unlike any other she has taken, and that more tests should be contextualized and theme based. The point of trialing a NS is to discover aspects of the test that do not make sense to NS, as a means to make sure there is nothing on the test that will further confuse a NNSs. The idea is that if there is something on the test that confuses NSs, than most likely it will also be troublesome for a NNS. The results were that the test was clear enough, because NS-B scored high, but that is all that can be determined by the score.

**NNS vs. NS Test Trials Reflection & Future Considerations:**

It is essential to explain the different types of test takers that the test Ben’s Story was trialed on, for future test trials results and actions to be clear, precise, and consistent. There was a total of nine Volunteer Trial Test Takers (VTTT). The nine VTTT consisted of NNS-students, NNS-non-students, NS-students, and NS-non-students, who all have different experiences, histories, and English language abilities. The problem that occurred, was that the results of all the different types of VTTTs’ were compared and contrasted against each other. As a result, some aspects of the test, like word choice, context, and tasks were altered.

This is a problem, because it is not justifiable to compare and contrast results that were all so very drastically different due to the variation in experiences, histories, and English language ability levels. Therefore the VTTT test results are all invalid, as well as any changes or assumptions made on behalf of those test results.

It is evident that there is a difference between the error patterns that occur in the results of tests taken by NSs versus tests taken by NNSs. This error pattern is further distinguished and established with other participating and influential aspects of the NS’s or NNS’s history, experience, and current learning status. For example, is the NS a student or not (learning status)? Is the NS currently receiving
some form of instruction in English that makes them a student? Is the NS not receiving instruction in English? If not, how long has it been since they attended any institution with English language instruction?

Furthermore, the same can be asked about the NNSs. Is the NNS a student or not? Has the NNS received training in English in the past? Is the NNS currently receiving English language instruction? The hypothesis is that these aspects of a VTTT affect the results of the tests. This is clearly visible in the results of the nine VTTT chosen to take Ben’s Story, because the backgrounds and scores are so different. The question remains, how could the different types of VTTT not play a role in the results of the tests when evaluated as whole? The difference in VTTT risks the trial test’s results accountability.

In the future, a group of VTTT should be chosen based on the similarities with the target test takers (TTT). For example, in the context of Ben’s Story, the TTT, are older, immigrant, and level 4 English language NNS-students. Therefore, the VTTTs should also be NNS-students with similar experiences, histories, and English language abilities. The similarities between the VTTT NNS and the TTT NNS will provide a reliable and consistent pattern of error that can be further studied to determine future developmental changes in a test. There can also be a group of VTTTs who are NSs, yet their scores cannot be compared and contrasted against the VTTT NNS, for the sake of making drastic changes to the test. Rather the VTTT NSs results will be accommodating in recognizing and discovering a pattern of error that can be used to inform decisions made on the test that benefit NNS TTTs.

Continue onto next page...
Target Test Taker Results

The results of the target test takers (TTT) were both negative and positive according to the scores (appendix b) and the Student Test Survey (STS) (appendix c). According to the scores, 12 out of 22 students did well, because they scored a 70 percent or above. That is roughly a little less than half of the total 22 students who took the Authentic Task Based Achievement Test (ATBAT) (Ben’s Story). Unfortunately, whether the language learners scored well or not is irrelevant at MCSA. The ESL-4 class at MCSA is less traditional and weekly or monthly test scores matter very little. Weekly and Monthly assessments are not required at McKinley (MCSA), because their forms assessment rely heavily on computer based proficiency placement tests. If the class was a bit more traditional and relied on continuous test scores to determine learner language ability, the results of the ATBAT (Ben’s Story) would prove positive, because most of the learners did well.

Score Analyses

According to the scores chart (appendix b), component part one and part three were the most difficult sections. Part one consists of reading and reading comprehension questions, and nine out of twenty-two student scored a 20 or above (appendix a, part 1). Part three consists of a vocabulary section, and five students out of twenty-two scored a 20 or above (appendix a, part 3). Surprisingly, fifteen out of twenty-two students scored a 20 or above in part four. Part four consists of a writing task which required learners to write a professional complaint letter (appendix a, part 4). Finally, part two proved the second easiest section based on the students’ scores (appendix a, part 2). Seventeen out of twenty-two students scored a twenty or above in part two, which consists of reading a dialogue, retaining written directions, and interpreting a map. Sadly, the last four students who scored 0 out of twenty-two students in section two, either did not have their glasses to interpret the map or did not read or understand all the instructions at all.
Scores were evaluated based on rubrics (see specifications, or appendix c for rubrics).

**Item Analyses**

According to the Item Analyses Chart (IAC) (appendix b), items a, f, and i were the easiest, because most of all the students (Ss) got it correct (check marks). Item b was the most difficult, because only five Ss got the item correct. The rest of the items, are fairly easy items, because some Ss got the items correct and some did not.

Along with the IAC chart, the IF chart portrays items a, f, and i, as the easiest, because of the percentages, 0.85, 0.75, and 0.70 are closest to 100% on a scale of 0% to 100%. The closer the percentage is to 0% the most difficult it is, and if the percentage is closer to 100%, then the item is easy. Items a,f, and i, placed closer to 100, making the items according to the chart and the percentages a lot easier. The rest of the items, c,d,e,g,h, and j, are fairly or somewhat easy, landing closer to 50% on the same scale from 0%-100%. The item that placed the most difficult, and that is closer to 0% is item b, which placed a 0.25%.

Furthermore, the ID chart exhibits the items of the vocabulary part of the test that discriminates between strong and weak learners. Items a & b are the least discriminating, because they scored a .2 on a scale from -1 and +1, with 0 in the middle. The results of the ID are items that are either discriminating or not, depending on how close the number is to 0. For example, items a, c, f, & j, are somewhat discriminating between strong and weak learners, because the items placed at 0.4, halfway between 0 and +1. The items that were the most discriminating are items d,g,h, and i., because the items placed at 0.6 or 0.8, which is closer to +1. Items d,g,h, and i placed closer to discriminating than any other item in the vocabulary part of the test. (Appendix b).
Reflection and Discussion
Language Assessment Concepts

Authentic Tasks & Performance-Based Assessment

The main concepts of the assessment project are authentic tasks and performance based assessment. A test that utilizes authentic tasks and content is necessary specifically for the target audience. Students at McKinley Community School for Adults (MCSA), are older, more experienced, and are interested in gaining skills to be able to perform outside of school. In order to have students really learn the material, connect to lessons and topics, and use the material outside of class, the tasks and content must be authentic, as in reflect real life tasks and situations. Newman, F. (1998) mentions that student’s also need more time to, “interpret documents, evaluate perspectives, theories and principles, and think for themselves,” in language learning, because often the work in language learning focuses too much on language forms and information retention, and not on information manipulation and utility (p. 2).

Authentic tasks in assessment allow for language learners to do just that, because the authentic tasks “consist of more than the ability to do well on an academic and traditional tests (Newman, 1991, p. 1). Authentic tasks in assessment contain real life and relevant tasks that learners can use outside in the real world. Authentic tasks in class and in assessment allow for language learning to become about purposeful learning versus “trivial and useless” learning (Newman, 1998, p.1).

Once a lesson and topic is relevant and purposeful to the language learners, language learners are more interested in language learning. Not only is “teaching and learning exciting,” at this point, but the achievement in authentic tasks is “significant and meaningful (Garran, 2008, p. 4)” Adding meaning, relevance, and purpose to the tasks in language learning and language assessment provides for
more positive wash back, and decreases fear and intimidation in language learning and language assessment.

Performance based assessment, “is assumed to support educational impact and learning,” and consist of more “thoughtful learning”, because language learners have the opportunity to process the information being learned through a performance, a demonstration, or a group project (Garran, 2008, p. 4). Language learners are also afforded “concurrent coaching” and consistent feedback (Miller and Archer, 2010, p. 5). Language learners need the opportunities to perform and process the target language with positive constructive criticism from both the teacher and peers in order to continue succeeding. Performance-based tasks and learning is especially critical with production skills. Newman and Wehlage (2003) mention that with performance-based language learning, “talking to learn and understand”, is a lot more powerful, than to simply talk for the sake of pronunciation, fact seeking, or defining. In performance-based language learning and assessment there can be, “considerable interaction about ideas of a topic,” and the possibilities and opportunities for more “higher order thinking, making distinctions, applying ideas, forming generalizations, and raising questions,” versus simply learning “facts, definitions and procedures (Newman and Wehlage 2003, p.4).”

Performance-based assessment also allow language learners to “demonstrate application of ideas, concepts, and principles,” of language learning (Garran, 2008, p.4). Performance-based teaching and assessment works well with traditional methods of instruction, for example, “class discussion, guided reading, writing assignments, note taking, and group learning” (Garran, 2008, p. 5). Unlike traditional methods being used in a language learning class, performance-based learning and assessment encourages and nurture the abstract and critical thinking skills in a language learner. Language learners should be able to, “manipulate information more readily, and think more creatively about content” (Garran, 2008, p. 5). Hence, the language learning process becomes an “experience” and not just an accumulation of classroom and lecture hours.
Components of the Test

Validity

Validity in assessment exists in multiple facets. According to Hughes (2003) there is “content validity, criterion-related validity, construct validity, validity in scoring, and face validity” (p.26-32). Validity in assessment means the difference between tests that assess what it was designed to assess versus a test that does not. Hughes (2003) says, “A test is said to have content validity if its content constitutes a representative sample of the language skills or structures with which it is meant to be concerned” (p.26). Validity in language assessment is extremely important, and it was a goal to make sure the instrument designed (see instrument) for the project was valid.

In the beginning, all four parts were designed separately to further evaluate and construct items using a testing technique like multiple choice (MC) or gap-fill (GF) that truly reflects face validity. Later, the test was evaluated as a whole to determine how well all four parts worked together, or against each other, and if the achievement test as a whole assesses the goal skills or structures. It was critical that the assessment design team focus on validity, not just for the sake of being valid, but because tests are already confusing to the target audience. Hughes (2003) expresses that a test does not successfully test what it aims to test will have “harmful wash back effect, because areas that are not tested, and tested correctly can become areas ignored in teaching and in learning” (p.27). Furthermore, administrating a test that did not have content validity, or face validity would have been catastrophic and inappropriate in attempting to measure the language learner’s targeted language learning skills. As a result, each section of the test was designed with simple task in mind, for students to complete, and was made sure it tested fulfilled the objective. Hughes (2003) states, “the greater a test content validity, the more likely it is to be an accurate measure of what it is supposed to measure” (p.27).
Context Issues

The context of the achievement test revolved around a real life situation and experience language learners would have to live through. Often with traditional tests there is little context or no context at all. Often such traditional tests do not allow for language learners to, “use their minds well,” and the assessment and the work required of the language learners to complete, has no “no meaning or value” (Newman, 2003, p.1). The context of “shopping” was relevant and interesting, because language learners shared their enthusiasm about learning how to communicate in a situation where a “refund” after shopping was necessary. Language learners also verbally expressed their interest in being able to write a professional complaint letter, because they felt it is important to express ones likes or dislikes of a situation.

The issue that arose with the context of shopping in the test, was making the context as authentic as possible through the use of a narrative that language learners could relate to. Performance based tests require language learners to, “demonstrate their knowledge in context of tasks,” yet still be “sensitive enough to determine language learners abilities to communicate” in the given context (Bailey, 1998, p. 209). Unfortunately, the instrument designers are not narrative or creative writers. The difficulties lied in constructing a narrative within the chosen context that was relevant, amusing, and entertaining, yet purposeful and useful. Making the narrative as relevant as possible was necessary to keep the target language learners interested and willing to participate. It is important that the language skills being test are “relevant but also practical”, to provide positive wash back (Bailey, 1998, p. 209). Language learners need to feel that they are learning a language they can use in a “practical way” (Bailey, 1998, p. 209). As a result, several revisions of the narrative were made, as well as revisions for the items for each part, yet it would have been more beneficial had the context of the achievement test been supported by a narrative that was designed by a professional. Nonetheless, other problems arose.
Issues of Test Development

Some of the issues in developing an Authentic Task Based Achievement Test (ATBAT) (Ben’s Story) was writing the communicative & comprehension questions, connecting all four components that utilized different testing techniques, as well as formulating a test that was friendly and comfortable, but neither too simple nor too complicated.

If the test were designed a bit more traditional, authentic tasks would not be a part of the test. If the test were a lot more traditional it would not have proven so difficult to design. Nonetheless, the test needed to be less traditional and task based, because of the TTT learning styles, and to originally test what the TTT actually learned in class. As a result, the goal was to use a similar questionnaire format Ss were accustomed to, yet it was a priority to test more than just reading comprehension, so the test questions became communicative. As a result, writing the questions were painful, because the questions, like the rest of the test needed to be simple. The cycle of revising the questions existed in the following manner. First, one of the group members wrote out simple reading comprehension questions. Then a second group member revised the questions to ask about better and specific parts of the story. Lastly, the last group member tried to formulate the questions to be more communicative. The results were a combustion of confusing, quizzative, and problematic questions, because each question could not provide the answer to the next. After a couple more revision cycles, the test was prepared for the VTTTs.

The second issue became connecting each part of the test, so that not only did the narrative make sense, but each part did also. The story had to be authentic, but it also had to follow a series of actions consistent to that of a real life process.

In addition, it was discovered after each part was made to connect to the test as a whole that each part provided the answers to the next part. This was a reality shock, because the test was designed to be administered as a whole. As a result, the executive decision was made to administer the test parts separately, at different
times. This was necessary to make sure the test recorded the TTTs full potential completing the tasks required of them in each part.

Not only did the items prove difficult to create, because they were based on a fictional narrative, but it was a struggle to design a test that is friendly. It is common for students, and other non-student personals to fear tests. The whole American society relies too heavily on the scores of test, when it has been proven that tests do not provide or reflect students’ abilities 100%.

Nonetheless, the test needed to be friendly and comfortable, not only to influence positive wash back, but because of the TTTs. As previously stated in the first paragraph under Results, tests matter very little to the TTT, and the school program. But it is essential to assess the TTT abilities, as a means to provide constructive feedback, and to better understand their strengths and weaknesses. The results as per STS was that the test was attractive, but the test was too long. Therefore, it had too many parts, and it took too long to complete it.

Although the test was not changed according to the results, this research has proven significant in testing of the TTTs at MCSA.
Future Inquiries

Goal
In the future it is a goal to embrace a Learner-Centered Language Learning method using task-based activities, as well as designing versatile authentic-task performance-based assessments.

Action Research Questions
Based on the practical experience in designing and implementing the assessment instrument and reviewing literature three questions that arose for action research in the future:

1. How can language assessments of all skills implement authentic-task performance-based assessments?
2. What are the strengths and draw backs of using authentic-task & performance-based assessment?
3. Can authentic-task & performance based assessment be used with all styles of English language learners?
References

Articles:

Books:
Appendices

Appendix a

Ben’s Story

Part 1: Buying Bread

Instructions: Please read this story about Ben and answer the questions using complete sentences.

On the way home, from a visit to a friend’s house, Ben visited the FoodCountry in Waipahu. After Ben had picked up a loaf of bread and some other groceries, he drove to his home in Honolulu. When he got home he went to put the bread away, but as he picked up the loaf of bread, chunks of bread fell through the plastic bag. Ben stopped and turned the bread over to see where the chunks of bread came from. He was shocked, because there were tiny bite marks on the plastic bag, and a hole on the bag the size of a quarter. There were also missing pieces of bread, and the bread looked like something had taken a bite out of it. Ben put the bread back into the FoodCountry bag and looked for his receipt. Once he had found his receipt, he put it in his wallet.

1. What do you think had happened to the bread?

2. Ben put a receipt in his wallet. What do you think he needed it for?

3. Would you eat Ben’s bread? Why/why not?

4. What do you think Ben will do next?

5. Has something similar happened to you or someone you know? What did you or that person do?
Part 2: Ben Receives A call from Anna

Instructions: Read the following dialogue
As Ben put away the other groceries his wife Anna calls.

Anna: “Hello honey, how are you?”
Ben: “I am well, but I went to the FoodCountry in Waipahu and bought bad bread.”
Anna: “What do you mean bad bread?”
Ben: “Well, the plastic bag has holes in it, the bread is also eaten, and chunks are falling out. It’s very bad.”
Anna: “Well honey, why don’t you go and return it?”
Ben: “But I bought it way out in Waipahu.”
Anna: “Why don’t you go to the FoodCountry in Kalihi?”
Ben: “I guess I can, where is that at? I have never been there before.”
Anna: “It is on N. School Street. In order to get there, go straight on Nuuanu Avenue towards the Foster Botanical Garden. Take a right on N School Street. Keep going straight for about 2 blocks, until you see the store on your right. The store is at the intersection of N School Street and Liliha Street. Oops! I’ve got to hang up now! Bye honey!
Ben: “Thanks honey, bye!!”

Instructions:

Ben now knows where the store is. He uses a map on the next page to look it up. First, find “Ben’s Location” on the map, and then follow Anna’s instructions. Draw the route on the map, and mark with an X where FoodCountry is located.
Part 3: Back to FoodCountry

Instructions: Please read each sentence and use words from the word bank to fill in the blanks. You do not have to use all the words.

1. Ben arrived at the FoodCountry in Honolulu. When he entered the store he needed to find a/an (a) _______________ to speak to.

2. He wanted to address the (b) ________________ with the bread, and (c) _______________ it.

3. Several other customers were trying to get their money back by asking for a (d) ________________, but they were complaining about many different types of (e) _______________, not just bread.

4. John the clerk asked Ben if he (f) _________________ the bread from FoodCountry.

5. Ben started looking for his (g) __________________, because he remembered that he had put the (h) _________________ in it, but unfortunately he couldn’t find it.

6. He told John the clerk that he had paid in (i) _________________, and wanted his money back.

7. John said he could only give Ben (j) __________________. This meant that Ben could not get his money back, but he could buy something else in the store for the same price as the bread.

After 10 minutes, Ben headed back home.

<table>
<thead>
<tr>
<th>return</th>
<th>receipt</th>
<th>clerk</th>
</tr>
</thead>
<tbody>
<tr>
<td>purchased</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>cash</th>
<th>payment</th>
<th>items</th>
<th>store</th>
</tr>
</thead>
<tbody>
<tr>
<td>credit</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| refund | wallet | bread | issue |
Ben was not happy about the return. He went into his office to write a business letter to FoodCountry. He looked their office up. It is at 94-1040 Waipio Uka St in Waipahu, HI 96797. He was very unhappy that the bread was ruined. It appeared that Foodcountry had had several other customers complain about similar issues. There were bite marks and several of their products seemed to have been eaten, so Ben wanted to send a letter to the FoodCountry manager in Waipahu. His name is Gerald Homes.

Instructions:
Imagine that you are Ben. Write a business letter to Gerald Homes reporting the issue with the bread in the business letter format that has been presented in class. Make sure you also include a request for what Gerald Homes should do (positive advice).
Appendix b

Scores Chart

<table>
<thead>
<tr>
<th>Scores</th>
<th>Column1</th>
<th>Column2</th>
<th>Column3</th>
<th>Column4</th>
<th>Column5</th>
<th>Column6</th>
<th>Total Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Part 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Part 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Part 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Part 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Item Analysis: Part 3 Vocabulary

<table>
<thead>
<tr>
<th>Item Analysis for Part 3: Vocabulary</th>
<th>a) clerk</th>
<th>b) issue</th>
<th>c) return</th>
<th>d) refund</th>
<th>e) items purchased</th>
<th>f) wallet</th>
<th>g) receipt</th>
<th>h) cash</th>
<th>i) store credit</th>
<th>IF</th>
<th>ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student A</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>a) 17/20=0.85</td>
<td>a) 8/8/2.55x20x0.4</td>
</tr>
<tr>
<td>Student B</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>b) 15/20=0.75</td>
<td>b) 1/1/2.55x20x0.2</td>
</tr>
<tr>
<td>Student C</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>c) 10/20=0.50</td>
<td>c) 4/4/2.55x20x0.4</td>
</tr>
<tr>
<td>Student D</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>d) 8/20=0.40</td>
<td>d) 4/4/2.55x20x0.8</td>
</tr>
<tr>
<td>Student E</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>e) 10/20=0.50</td>
<td>e) 2/2/2.55x20x0.2</td>
</tr>
<tr>
<td>Student F</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>f) 14/20=0.70</td>
<td>f) 5/5/2.55x20x0.4</td>
</tr>
<tr>
<td>Student G</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>g) 12/20=0.60</td>
<td>g) 5/5/2.55x20x0.3</td>
</tr>
<tr>
<td>Student H</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>h) 10/20=0.50</td>
<td>h) 5/5/2.55x20x0.6</td>
</tr>
<tr>
<td>Student I</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>i) 15/20=0.75</td>
<td>i) 5/5/2.55x20x0.6</td>
</tr>
<tr>
<td>Student J</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>j) 15/20=0.75</td>
<td>j) 5/5/2.55x20x0.6</td>
</tr>
<tr>
<td>Student K</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>k) 15/20=0.75</td>
<td>k) 5/5/2.55x20x0.6</td>
</tr>
<tr>
<td>Student L</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>l) 13/20=0.65</td>
<td>l) 4/4/2.55x20x0.4</td>
</tr>
<tr>
<td>Student M</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>Easy Items</td>
<td>Least Discriminating</td>
</tr>
<tr>
<td>Student N</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>Fairly Easy Item</td>
<td>Somewhat Discriminating</td>
</tr>
<tr>
<td>Student O</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>Least Easy Item</td>
<td>Highly Discriminating</td>
</tr>
<tr>
<td>Student P</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>Least Discriminating</td>
<td></td>
</tr>
<tr>
<td>Student Q</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>Fairly Easy Item</td>
<td></td>
</tr>
<tr>
<td>Student R</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>Least Discriminating</td>
<td></td>
</tr>
<tr>
<td>Student S</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>Fairly Easy Item</td>
<td></td>
</tr>
<tr>
<td>Student T</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>Least Discriminating</td>
<td></td>
</tr>
</tbody>
</table>
Appendix c

**Rubrics** for Authentic Task Based Achievement Test (ATBAT) = Ben’s Story

**Part 1**

Rubric

- (5) Complete sentence with relevant information
- (2.5) Incomplete sentence with relevant information
- (0) Complete or incomplete sentence with irrelevant information

**Answer Key**

6. Student proposes that something capable of leaving bite-marks has taken a bite out of the bread.
7. Student proposes that Ben needs the receipt to return the bread.
8. Students explain why or why they wouldn’t eat Ben’s bread backing it up with content in the reading.
9. Student envisions that Ben does something that would be likely given the context
10. Student answered the question (yes or no) and provided a description if applicable

**Part 2**

Answer key

- (5) The student placed the X correctly
- (5) The student has drawn a route that leads to the X
- (5) The student’s route goes through Nuuanu Avenue towards Foster Botanical Garden
- (5) The student’s route goes through North School Street
- (5) The student’s route ends after following North School Street for two blocks

**Part 3**

Answer key

i. (2.5) Clerk
j. (2.5) Issue
k. (2.5) Return
l. (2.5) Refund
m. (2.5) Items
n. (2.5) Purchased
o. (2.5) Wallet
p. (2.5) Receipt
q. (2.5) Cash
Appendix d

Student Test Survey (STS)

A. Instructions: Please circle the best answer. You can circle more than one answer.

1. Which part of Ben’s Story is easy?
   a. Part 1: Reading and Questions
   b. Part 2: Dialogue and Map
   c. Vocabulary Multiple Choice
   d. Business/ Complaint Letter?

2. Which part of Ben’s Story is difficult?
   a. Part 1: Reading and Questions
   b. Part 2: Dialogue and Map
   c. Vocabulary Multiple Choice
   d. Business/ Complaint Letter?

B. Instructions: Please check off the best answer.

3. Was the test interesting? (Yes _____ No _____)
4. Can you relate to Ben’s Story? (Yes____ No _____)
5. Was the test too short or too long? (Too short ____ Too long _____)
6. Were the instructions clear or unclear? (clear _____ unclear____)

C. Instructions: Please answer each questions as best you can.

7. What did you like about the test? Please give two reasons why.

________________________________________________________________________________________________________

________________________________________________________________________________________________________

________________________________________________________________________________________________________

8. What didn’t you like about the test? Please give two or more reasons why.

________________________________________________________________________________________________________

________________________________________________________________________________________________________

________________________________________________________________________________________________________

9. What positive advice (suggestions) would you give to the teachers who designed this test?

________________________________________________________________________________________________________

________________________________________________________________________________________________________

________________________________________________________________________________________________________