Amendment No. 2
to
Contract No. NA13000067
for
Language Proficiency Exam
between
Berlitz Language Inc.
and the
City of Austin

1.0 The City hereby exercises the extension option for the above-referenced contract. Effective February 11, 2016 term for the extension option will be February 11, 2016 to February 10, 2019, no options remain.

2.0 The total contract amount is increased by $112,500.00 for the extension option period. The total Contract authorization is recapped below:

Term	Action Amount	Total Contract Amount
Basic Term: 02/11/13 – 02/10/16	$112,500.00	$112,500.00
Amendment No. 1: Add APD and AFD as authorized department 2/11/2013 – 2/10/2016	$0.00	$112,500.00
Amendment No. 2: Option 1 02/11/16 – 02/10/19	$112,500.00	$225,000.00

3.0 MBE/WBE goals were not established for this contract.

4.0 By signing this Amendment the Contractor certifies that the Contractor and its principals are not currently suspended or debarred from doing business with the Federal Government, as indicated by the General Services Administration (GSA) List of Parties Excluded from Federal Procurement and Non-Procurement Programs, the State of Texas, or the City of Austin.

5.0 All other terms and conditions remain the same.

BY THE SIGNATURES affixed below, this Amendment is hereby incorporated into and made a part of the above-referenced contract.

Signature & Date: 2/4/16
Printed Name:
Authorized Representative

Signature & Date: Linell Goodin-Brown
Linell Goodin-Brown, Contract Compliance Supervisor 2-26-16
City of Austin
Purchasing Office

Berlitz Language Inc
7000 North MoPac Expwy, Suite 200
Austin, TX 78731

Amendment No. 1
to
Contract No.NA130000067
for
Language Proficency Exam
between
Berlitz Languages Inc.
and the
City of Austin

1.0 The City hereby amends the referenced contract to addthe Austin Poice Department (APD) and Austin Fire Department (AFD) as authorized departments to the contract for pre-employment and promotional testing purposes.

2.0 The total contract authorization is recapped below:

Term	Action Amount	Total Contract Amount
Initial Term: 2/11/2013 – 2/10/2016	$37,500.00	$37,500.00
Amendment No. 1: Add APD and AFD as authorized deparments. 2/11/2013 – 2/10/2016	$0	$37,500.00

3.0 MBE/WBE goals do not apply to this contract.

4.0 By signing this Amendment the Contractor certifies that the vendor and its principals are not currently suspended or debarred from doing business with the Federal Government, as indicated by the GSA List of Parties Excluded from Federal Procurement and Non-Procurement Programs, the State of Texas, or the City of Austin.

5.0 All other terms and conditions remain the same.

BY THE SIGNATURES affixed below, this amendment is hereby incorporated into and made a part of the above-referenced contract.

Sign/Date: 3/14/2014

Printed Name: Paul H. Weinstein, VP
Authorized Representative

Sign/Date: 3/17/14

Lynn Rich
Buyer II
City of Austin
Purchasing Office
124 W. 8th Street, Ste. 330
Austin, Texas 78701

City of Austin

Financial and Administrative Services Department
Purchasing Office
P. O. Box 1088, Austin, TX 78767 (512) 974-2500

2/4/2013

Kristen Erdem
Berlitz Languages, Inc.
7000 North MoPac STE 200
Austin, TX 78731

Re: Solicitation No. JSD0129

Dear Ms. Erdem,

The Purchasing Office has approved the execution of a contract with your company for the above-referenced item as follows:

Responsible Department:	HR
Department Contact Person:	Holly Moyer
Department Contact Email Address:	Holly.moyer@ci.austin.tx.us
Department Contact Telephone:	(512) 974-3216
Project Name:	Language Proficiency Exam Services
Contractor Name:	Berlitz Languages Inc
Contract Number:	MA 5800 NA130000067
Contract Period:	2/11/2013 –2/10/2016
Dollar Amount	NTE $112,500
Extension Options:	1 -36 month
Requisition Number:	RQM-5800-12083000531
Solicitation Number:	JSD0129
Agenda Item Number:	22
Council Approval Date:	1/31/2013

Thank you for your interest in doing business with the City of Austin. If you have any questions regarding this contract, please contact me at (512) 974-2651

Sincerely,

Jeff Dilbert, MSIM, CPSM, CPSD, C.P.M., A.P.P.
Corporate Purchasing Manager
Purchasing Office
Finance and Administrative Services Department
Enclosure

CONTRACT BETWEEN THE CITY OF AUSTIN ("City")
AND
Berlitz Languages, Inc. ("Contractor")
for
Language Proficiency Exam Services
MA 5800 NA130000067

The City accepts the Contractor's Offer (as referenced in Section 1.1.3 below) for the above requirement and enters into the following Contract.

This Contract is between Berlitz Languages, Inc. having offices at Austin, TX 78731 and the City, a home-rule municipality incorporated by the State of Texas, and is effective as of the date executed by the City ("Effective Date").

Capitalized terms used but not defined herein have the meanings given them in Solicitation Number RFP JSD0129.

1.1 **This Contract is composed of the following documents:**

1.1.1 This Contract

1.1.2 The City's Solicitation, Request for Proposal (RFP), JSD0129 including all documents incorporated by reference

1.1.3 Berlitz Languages, Inc. Offer, dated 10/12/2012, including subsequent clarifications

1.2 **Order of Precedence.** Any inconsistency or conflict in the Contract documents shall be resolved by giving precedence in the following order:

1.2.1 This Contract

1.2.2 The City's Solicitation as referenced in Section 1.1.2, including all documents incorporated by reference

1.2.3 The Contractor's Offer as referenced in Section 1.1.3, including subsequent clarifications

1.3 **Quantity of Work.** There is no guaranteed quantity of work for the period of the Contract and there are no minimum order quantities. Work will be on an as needed basis as specified by the City for each Delivery Order.

1.4 **Term of Contract.** The Contract will be in effect for an initial term of thirty-six (36) months and may be extended thereafter for up to one (1) thirty-six (36) month extension option(s), subject to the approval of the Contractor and the City Purchasing Officer or his designee. See the Term of Contract provision in Section 0400 for additional Contract requirements.

1.5 **Compensation.** The Contractor shall be paid a total Not-to-Exceed amount of $112,500 for the initial Contract term and $112,500 for each extension option as indicated in the Bid Sheet, IFB Section 0600. Payment shall be made upon successful completion of services or delivery of goods as outlined in each individual Delivery Order.

This Contract (including any Exhibits) constitutes the entire agreement of the parties regarding the subject matter of this Contract and supersedes all prior and contemporaneous agreements and understandings, whether written or oral, relating to such subject matter. This Contract may be altered, amended, or modified only by a written instrument signed by the duly authorized representatives of both parties.

In witness whereof, the City has caused a duly authorized representative to execute this Contract on the date set forth below.

CITY OF AUSTIN

Printed Name of Authorized Person:	Jeff Dilbert
Signature:	
Title:	Corporate Purchasing Manager
Date	2/4/2013

CITY OF AUSTIN, TEXAS
Purchasing Office
REQUEST FOR PROPOSAL (RFP)
Offer Sheet

SOLICITATION NO.: JSD0129

DATE ISSUED: 9/17/2012

REQUISITION NO.: RQM 1206300531

COMMODITY CODE: 90783

FOR CONTRACTUAL AND TECHNICAL ISSUES CONTACT

Jeff Dilbert
Corporate Purchasing Manager
Phone: (512) 974-2021

COMMODITY/SERVICE DESCRIPTION: Language Proficiency Exam Services

PRE-PROPOSAL CONFERENCE TIME AND DATE: N/A

LOCATION: N/A

PROPOSAL DUE PRIOR TO: 2:00p.m. on 10/12/2012

COMPLIANCE PLAN DUE PRIOR TO: N/A

PROPOSAL CLOSING TIME AND DATE: 2:00p.m. on 10/12/2012

LOCATION: MUNICIPAL BUILDING, 124 W 8th STREET
RM 310, AUSTIN, TEXAS 78701

When submitting a sealed Offer and/or Compliance Plan, use the proper address for the type of service desired, as shown below.

P.O. Address for US Mail	Street Address for Hand Delivery or Courier Service
City of Austin	City of Austin, Purchasing Office
Purchasing Office	Municipal Building
P.O. Box 1088	124 W 8th Street, Rm 310
Austin, Texas 78767-8845	Austin, Texas 78701
	Reception Phone: (512) 974-2500

Offers (including Compliance Plans) that are not submitted in a sealed envelope or container will not be considered.

SUBMIT 1 ORIGINAL AND 4 SIGNED COPIES OF PROPOSAL

OFFER SUBMITTED BY

By the signature below, I certify that I have submitted a binding offer.

Kristen Erdem

Signature of Person Authorized to Sign Offer

Kristen Erdem, Business Development Manager for Texas and Oklahoma

Signer's Name and Title (please print or type)

FEDERAL TAX ID NO. [redacted]

Date: 10/12/12

Company Name: Berlitz Languages, Inc.

Address: 7000 North McPac Suite 200

City, State, Zip Code: Austin, TX 78731

Phone No. (512) 674-1957

Fax No. ( )

Email Address: kristen.erdem@berlitz.us

Request for Proposals

PURCHASING OFFICE
CITY OF AUSTIN, TEXAS

Language Proficiency Exam Services

RFP No.: JSD0129 **Addendum No. 1** **Date of Addendum: October 12, 2012**

1.0 **The following clarifications are provided in response to questions from potential respondents:**

1. In 2.2, the City of Austin requires that results be reported as Pass/Fail – has a passing rating on any established proficiency scale (the ACTFL scale or the Interagency Language Roundtable scale, for example) been established? If so, what is the passing rating required? No If not, is the bidder expected to propose a method of establishing pass/fail levels for particular jobs and include appropriate pricing information in Attachment B? Yes
2. IN 2.8, the City requires a fixed testing site. Would the City be open to tests taking place in City offices, and being proctored by locally-approved City HR staff or line-managers. No
3. In 4.1 the City requires that testing be done face-to-face – would the City be open to proposals using telephonic speaking and web-based writing tests instead? Yes
4. In 5.0, c., the City requires resumes of examiners – would the City be open to general descriptions of the qualifications and certification procedures ensuring that skilled, certified testers are utilized, as it is at this stage impossible to identify the individual testers who would provide service? Yes

2.0 **The proposal due date and time has been extended to October 19, 2012 @ 2:00 P.M.**

3.0 All other terms and conditions will remain the same

BY THE SIGNATURES affixed below, Addendum No. 1 is hereby incorporated and made a part of the above-referenced Request for Proposal.

APPROVED BY: Steve Cocke

Steve Cocke, Buyer II
Finance and Administrative Services Department

ACKNOWLEDGED BY	Kristen Erdem Berlitz Languages, Inc	Kristen Erdem	10/18/12
	Bidder	Authorized Signature	Date

RETURN ONE (1) COPY TO THE PURCHASING OFFICE, CITY OF AUSTIN, TEXAS PRIOR TO BID OPENING OR WITH YOUR BID. FAILURE TO DO SO MAY CONSTITUTE GROUNDS FOR REJECTION OF YOUR BID.

Request for Proposals

PURCHASING OFFICE
CITY OF AUSTIN, TEXAS

Language Proficiency Exam Services

RFP No.: JSD0129 **Addendum No. 2** **Date of Addendum: October 18, 2012**

1.0 The following clarifications are provided in response to questions from potential respondents:

1. Should we include all customization work in the price per examinee? Yes
2. Please tell us the expected volume of testing for each language other than Spanish? Unknown at this time
3. Can you please provide more detail on the scope of the customization to be requested? Customization will be related to the function of the position. For example if the position is in the medical field, we would request testing to include testing of medical terminology.
4. In the RFP, you ask for customization to be completed within ten days. Is it acceptable to count the ten business days from the date that we receive details of the request for customization and finishing on presentation of the revised items to the client? Yes, ten business days.
5. When are the tests expected to be operational? February 2013.

2.0 All other terms and conditions will remain the same.

BY THE SIGNATURES affixed below, Addendum No. 2 is hereby incorporated and made a part of the above-referenced Request for Proposal

APPROVED BY: Steve Cocke

Steve Cocke, Buyer II
Finance and Administrative Services Department

ACKNOWLEDGED BY:	Kristen Erdem Berlitz Languages Inc.	Kristen Erdem	10 18 12
	Bidder	Authorized Signature	Date

RETURN ONE (1) COPY TO THE PURCHASING OFFICE, CITY OF AUSTIN, TEXAS PRIOR TO BID OPENING OR WITH YOUR BID. FAILURE TO DO SO MAY CONSTITUTE GROUNDS FOR REJECTION OF YOUR BID.

Berlitz® Testing

Response to RFP No. JSD0129

Language Proficiency Exam Services

October 18, 2012

Submitted to:

Jeff Dilbert
Corporate Purchasing Manager
City of Austin, Purchasing Office
Municipal Building
124 W. 8th Street, Room 310
Austin, TX 78701
Phone: (512) 974-2021

Berlitz point of contact:

Kristen Erdem
Business Development Manager for Texas and Oklahoma
Berlitz Languages, Inc.
7000 North MoPac
Suite 200
Austin, TX 78731
Phone: (512) 674-1957
Email: Kristen.erdem@berlitz.us

TABLE OF CONTENTS

EXHIBITS

EXECUTIVE SUMMARY

Berlitz is pleased to respond to RFP JSD0129 by the City of Austin and would like to be considered to provide the language proficiency testing services needed to qualify city employees for its bilingual pay program.

Berlitz' existing testing products, its ability to customize them, as well as it its stability and breadth of experience in the testing field around the world and in Austin, Texas, qualifies it well for this testing program. We can deliver face to face service at our company location in Austin that you require with test products that have proven valid an reliable over time. We can deliver testing for all of the languages you require.

In this proposal, Berlitz proposes to deliver the Berlitz Test of Speaking Skills and the Berlitz Test of Writing Skills to employees of the City of Austin. The Test of Speaking Skills is an oral proficiency interview designed to reliably measure the speaking ability of employees. The test has been in existence for over 5 years and is administered by well-trained and supported human-raters. The Berlitz Test of Writing skills evaluates an employee's ability to write clearly and effectively in the most common work-oriented formats: e-mail, memo, and report writing. Both the speaking and the writing tests as well as the system for rating them have proven over the years to provide our customers consistently good information about the language skills of employees so that they can make higher stakes personnel decisions.

In order to provide the city of Austin with the benefit of the most experienced and calibrated raters, the Berlitz Test of Speaking Skills will be delivered over the phone with the examinee sitting at the Berlitz Language Center proctored by Berlitz Employees. For the writing test, we will proctor all tests at the Berlitz Language Center in Austin as well.

Second language Testing, (SLTI), a subsidiary of Berlitz that provides custom test development for government clients, will work with the City of Austin to provide the expert consulting on establishing cut scores (pass/fail scores), and the customization needed to provide you information not only on the proficiency level of employees, but also on their ability to use these skills in a relevant legal or medical context.

The ability to provide local and relevant testing services backed by the resources and experience of Berlitz and SLTI enables us to provide the quality of products, service, and stability that the City of Austin needs for its testing program.

PART 1 – Business Organization

Berlitz Corporation is a global Education company with headquarters in Princeton, New Jersey and Tokyo, Japan. The company was founded in 1878 by Maximilian D. Berlitz in Providence, Rhode Island. Berlitz Corporation is a member of Benesse Group, with more than 550 company-owned and franchised locations in over 70 countries offering instruction, assessment, cultural consulting and global leadership for individuals, businesses, government agencies, and non-profit organizations in more than 50 languages. The local Berlitz Austin location is located at 7000 North MoPac Expwy, Suite 200, Austin, Texas 78733. The Berlitz Center in Austin will be responsible for proctoring all tests and Kristen Erdem, our Director of Business Development in Austin will be the point of contact for all contract issues.

Berlitz's test development company, Second Language Testing, Inc. (SLTI) is based in Rockville MD. SLTI specializes in development of language proficiency tests and test translation and/or adaptation. SLTI will be responsible for any customization or Test Development work needed for this project, supports the Testing Administration Center in its product portfolio and consults closely with our office in Austin, Texas and other Berlitz centers thoughout the United States. SLTI has a staff of approximately 30 full-time employees, about a dozen part-time and temporary employees, and maintains a worldwide network of consultants, including testing specialists and psychometricians, as well as translators and language specialists for over 50 languages.

Second Language Testing, Inc. (SLTI) is the first company of its kind—focusing solely on test development and the translation of tests of language skills. SLTI works with a wide range of clients, ranging from standardized assessments for U.S. State Departments of Education, high-stakes assessments for college and university admissions, to end-of-unit assessments. SLTI also develops tests to certify government and military linguists and to credential courtroom and medical interpreters.

This base of experience and breadth of knowledge gives SLTI an advantage in dealing with the issues that may arise in any discussion of language proficiency tests. Further, SLTI's project managers are well versed in the procedures associated with large-scale test development projects, and there is a continual focus on developing and maintaining strict timetables for deliverables without compromising product quality.

The past accomplishments of SLTI should be viewed as an indication of the organization's potential to carry out the requirements of any proposed project. The background and skill of SLTI employees and consultants ensures that the final products of every project are of high quality that will withstand the tests of time and operational implementation.

Berlitz' test administration center is headquarted in Woodland Hills, CA and is responsible for recruiting, training, and maintaining qualified raters in over 40 languages. They schedule tests, report scores, and maintain the calibration of the rater pool so that our clients get consistently reliable results.

PART 2 – System Concept and Solution

The City of Austin requires a company to administer speaking proficiency tests for Spanish, French, German, Italian, Portuguese, Arabic, and Thai and writing proficiency tests for Spanish, French, German, Italian, Portuguese, Arabic, and Thai. Berlitz proposes the use of the Berlitz Test of Speaking Skills and the Berlitz Test of Writing Skills for all candidates.

We propose that both tests be proctored at the Berlitz Language Center in Austin, TX. The speaking test will be delivered over the phone through our testing center in Woodland Hills, CA by our existing, well-trained, and calibrated rater pool. The Berlitz test of Writing Skills will be proctored at the Austin Berlitz Center as well.

Both the speaking and the writing tests will be scored according to the Common European Framework of Reference (CEFR). The CEFR is a six point scale ranging from A1 to C2 and is the most widely accepted standard for proficiency around the world. See Appendix A for detailed descriptors of the scale. Berlitz Testing raters are trained and calibrated to this scale.

SLTI can work with the City of Austin to provide some customized components to both the Test of Speaking Skills and to the Test of Writing Skills. We understand that the City of Austin requires a ten day turnaround for customization work. In ten days, we would be able to develop additional prompts, pilot them with 1-2 examinees and revise them based upon the feedback received. In order to achieve this deadline, we would ask the City of Austin to provide us with 5 examples of relevant tasks each for writing and speaking that are consistent with the type and level of work that these employees would be asked to handle appropriately in their roles in the City of Austin. SLTI Test developers will use this information, in combination with other information that we have, as the basis for constructing additional writing prompts and speaking scenarios that would enable examinees to demonstrate their language proficiency in those relevant contexts. In 10 days, we are not going to be able to do the field testing of prompts that would provide the quantitative data that would enable us to provide the basis for a strong validity argument of any customization performed. Given this, we recommend utilizing the existing tests with minor customization. The existing tests will provide the basis for the validity argument and the customized pieces will provide us with closer look at language use in the City of Austin's context.

Upon contract award, Berlitz Testing will meet with City of Austin representatives to establish its customization needs as well as the cut scores on the CEFR. Any examinee falling below the cut score will fail the test and any examinee above the cut score will pass. Local Austin Staff will meet face to face and SLTI staff will join over the phone to avoid the inclusion of any the travel costs for the City of Austin.

Pass/Fail results will be delivered within 5 business of the exam date via e-mail to the City of Austin, Human Resources Compensation Division for each eligible employee. Please see Appendix C for an example of a pass/fail score report for the City of Austin. Berlitz Testing will provide appropriate experts to defend the reliability and validity of its tests and, if requested to do so provide it through expert testimony in court or administrative proceedings.

Berlitz Testing will provide information to the examinees including examination guidelines, notification of scheduling processes and policies, as well as confirmation of testing dates directly to the eligible employees and will submit a monthly program report to the City of Austin. Please see the Appendices E, G, and H for examples of these notifications.

Languages other than Spanish:

Berlitz currently supports all of the languages required by the City of Austin for its standard Test of Speaking Skills and its Standard Test of Writing Skills and can accommodate tests for those languages. We can, therefore, provide the same level of service for those languages as for Spanish. Since we are not able to ascertain the volume for customization of these languages, we must include customization as a separate charge for those languages. For any additional language required as part of this contract outside of our current available languages, Berlitz can usually accommodate you, but will charge a language initiation fee and will need 1 month notice. For a complete list of languages available for each of the Berlitz Test of Speaking Skills and the Berlitz Test of Writing Skills, please see Appendix N.

PART 3 – Program

Success in the implementation of our testing solution for the City of Austin will be ensured through good communication and adherence to well-established standards in the delivery and development of language tests. We follow a tested workflow for all of our testing projects, as described below.

3.1 Project Kick-Off and Needs Analysis

Timing:	Immediately after contract signing
Attendees:	*In Austin* City of Austin Human Resources Staff Kristen Erdem, Berlitz Director of Business Development Austin Claudia Diaz, Austin Language Center Manager *On the phone* Tom Godfrey, Director of Operations, NA Testing Alexia Nguyen, Manager, Testing Center, Woodland Hills CA Lauren Kennedy, Director of Test Development SLTI

Immediately following contract signing, Berlitz will meet with City of Austin Human Resource or other officials who can provide detailed information about the roles and types of tasks each examinee is expected to perform in the specified languages. If possible, they should bring with them examples of the types of work product that each employee is expected to produce or the types of tasks employees are expected to perform.

The agenda of the meeting will include the following plus any additions from the City of Austin

Introductions including roles and responsibilities	Kristen Erdem
Objectives	Kristen Erdem
Explanation of our program	Alexia Nguyen
The Common European Framework of Reference (CEFR)	Lauren Kennedy
Establishing cut scores on the CEFR	Lauren Kennedy and City of Austin Staff
Customization needs analysis	Lauren Kennedy
Scheduling and reporting Procedures	Claudia Diaz
Next steps and follow up	Kristen Erdem
Meeting Close	

3.2 Test Customization: Ten Days

Item Development:

Immediately following the meeting, Ms. Kenneedy will meet with her test developmet team to present the customization needs for the City of Austin. They will break the work down into discrete tasks and assign them to SLTI item writers with the relevant background to perform the tasks. Once initial drafts are complete, items will go through a review and revision process, which will continue until such time as we are satisfied that we have items that meet the client specificactions and that are consistent with industry standards for test development.

Pilot Testing

Once the items are created and reviewed, SLTI staff will pilot the items with two people who are similar to the profile of the candidates for roles at the City of Austin. In the piloting phase, we will catch any infelicities in item construction and ensure that examinees are engaging with the items as intended. Depending upon the results of pilot testing, items may go through an additional review and editing cycle.

3.3 Client Review

SLTI will present the customization work to the City of Austin for review and approval at a meeting to take place on the phone. SLTI will share its work, answer questions and listen to feedback from the client. At this point, we would be looking for the the City of Austin to approve the work so that we can move on to the operational phase.

3.4 Test Translation

Once customization items are approved by the client and it is clear what additional languages may be needed, SLTI's Test Translation Team will begin the task of adapting the tasks to additional languages. The Test Translation Team is made up of individuals with a background in both translation and in test development and so are able to ensure that when test items are translated, they appear natural and have the intended content and level characteristics as the original items. They will not simply be translations of items.

3.5 Rater Training: 2 days

Once the items have been approved by the client, SLTI staff will provide the raters orientation and training on the delivery of speaking scenarios and the scoring of tasks including examples of what to look for in passing tests.

3.6 Communication

The Berlitz Testing Center in conjunction with the Austin Learning Center will provide the City of Austin HR Department will an e-mail message that can be sent by the City of Austin or by Berlitz with detailed information about the test, scheduling procedures as well as the the location of the Austin Berlitz Center, (See Appendices B, D, and E).

3.7 At the Berlitz Center

On the day of the scheduled test, examinees will come to the Berlitz Center in Austin. They will be greeted by LC staff who will orient them to the location including showing them locations of facilities. The staff member will check the ID of the person, make a copy of their picture ID and hold the examinee's phone for the duration of the test. They will then be escorted to the room where they will take the test with a telephone. The proctor will dial the number of the testing center and introduce the examinee to the rater. The rater will make the examinee feel comfortable and will then request permission of the examinee to record the exam. Upon receipt of permission of the examinee to record the exam, the Berlitz Test of Speaking skills will commence. After the speaking test is complete the examinee will be provided the Berlitz Test of Writing skills and will be given 60 minutes to complete it. At the end of the 60 minutes, the proctor will collect all test papers and thank the examinee for coming.

Berlitz maintains a facility with 4 classrooms in the building that we rent, but are able to add additional spaces on a temporary basis (daily or hourly) with phones with a days notice should we have the need. We moved our center down the block to this facility specifically so that we could flexibly scale to meet our client needs. (See Appendices I and J)

3.8 Operational Testing: Berlitz Test of Speaking Skills

The Berlitz Test of Speaking Skills (BTSS) is an individually administered telephone conversation implemented by a trained Berlitz rater. In this Rated Conversation, the rater and the examinee engage in a conversation both about general topics and topics related to business and the workplace. The rater sets up a context for the conversation and asks questions that allow examinees to demonstrate their oral proficiency of the target language. The focus of the *Rated* Conversation is to engage the examinee in a conversation about topics related to the professional environment. All topics may be interchanged between levels. Raters change the way they ask a question in order to elicit the "expectation" at that level.

The smooth transition and progression from familiar topics to more challenging topics during the test process provides opportunities for the examinees to display their range of competence in managing the conversation effectively. The flexibility of the test is also displayed in the way raters focus on examinees' contexts and responses during the test. This allows raters to proceed with increasingly complex questions on a particular topic. Because of its flexible nature, the Test of Speaking Skills is not a scripted test.

The following list provides some topics used in the Rated Conversation:

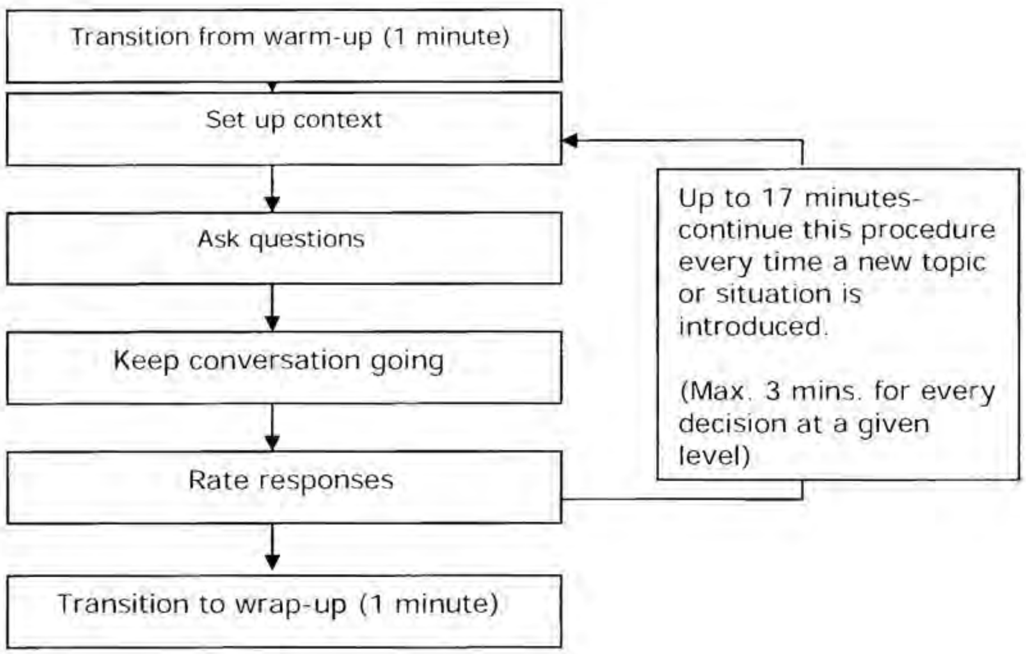
General topics
-Technology
-Travel
-Transportation
-Shopping
-Leisure
-Health

Professional Topics
-Work environment

-Tasks and responsibilities
-Customer service
-Handling customer inquiries
-Describing/advertising services and products
-Attending business meetings

Figure 1 illustrates the Rated Conversation Procedure.

Figure 1: Rated Conversation Procedure



BTSS Scoring Guide

The test assesses the examinee's oral proficiency in **four key areas** for each level of the CEFR scale: Linguistic Range, Grammatical Accuracy, Fluency, and Phonological Control.

Linguistic Range refers to how well speakers can express themselves using sentence patterns and forms, and idiomatic expressions that are appropriate in oral communications. Range also refers to the aspects of style, tone, and lexical choice.

Grammatical Accuracy may be defined as knowledge of, and ability to use, the grammatical structures of a language. Grammatical organization involves such styles and usages as word order, part of speech, and prepositions.

Fluency as a factor which determines the functional success of the examinees refers to the ability to articulate and manage conversation smoothly.

Phonological Control refers to the ability to produce appropriate intonation, variation in phonological phonemes (/t/ vs. /d/; /r/ vs./l/), correct sentence-stress placement to express finer shades of meaning, and production of clear intelligible pronunciation.

This Scoring Rubric is based on the CEFR Rating Scale and holistic descriptors. It rates examinees from level A1 (the lowest) to C2 (the highest).

More information regarding the BTSS can be found in Exhibits A – C.

3.9 Operational Testing: The Berlitz Test of Writing Skills

The Berlitz Test of Writing Skill (BTWS) assesses second language writing skills and was developed through the collaborative efforts of Berlitz International, Inc. and the University of Illinois at Urbana-Champaign. The test was designed to assess business and professional use of English writing in a workplace setting. This test is intended to be used by global organizations and government agencies in their recruitment, staff training, and benchmarking processes.

The BTWS includes three tasks that were chosen based upon research as the most relevant and authentic in a business context. The three chosen tasks are: write an e-mail, a memo, and a report respectively.

Examinees will be assessed based upon the following:

Task Fulfillment – We assess the examinee's understanding of a given input and their ability to meet the content requirements of the response including the ability to use a style appropriate to the audience.

Idea Organization – We assess the examinee's ability to organize text using cohesive devices and to plan text following the conventional structure of an essay including an introduction, body, and conclusion. Also assessed will be an examinees ability to group and present ideas with supporting information.

Language Use – We assess the examinee's grammatical and syntactic proficiency, vocabulary use and spelling.

BTWS rating scale

Operational Scoring Method

Two scoring methods, namely analytic scores and an average score that serves as a holistic score, are adopted for rating examinee essays in the operational scoring system. In the analytic scoring, raters assign three different scores per essay according to the three assessment criteria in the BTWS rating scale. Subsequently, a holistic score is calculated by averaging three analytic scores. Raters do not mark a holistic score independently.

Figure 2: Summary of the BTWS task attributes

Task Considerations	Task 1: Write an Email	Task 2: Write a Memo	Task 3: Write a Report
1. Communicative function	Request	Announce	Explain
2. Purpose	Convey objective information	Synthesize given information	Evaluate given information
3. Cognitive demand	Knowledge (lowest)	Analysis and synthesis (intermediate)	Analysis, synthesis and evaluation (highest)
4. Audience	Colleagues	Managers	Clients
5. Prompt input	E-mail (written)	Form (written)	Chart and graph (graphic)
6. Response length	Minimum 50 words	Minimum 120 words	Minimum 200 words
7. Test taking time	12 minutes	18 minutes	30 minutes
8. Expected formality		Less formal -- ---More formal	
9. Expected level of difficulty		Easy -- ---Difficult	

More information regarding the BTSS can be found in Exhibits D – F.

3.10 Reporting

Once tests are scored and reviewed, Berlitz Testing will send the score report, (see Appendix C) to the designated address at the City of Austin within 5 days of the test date. All test papers, score reports, and audio recordings of speaking tests will be maintained at the testing center for a minimum of 3 years.

Monthly summary reports with the results of all examinees will be provided by the 15th of each month including the scores and examinees. It will be accompanied by an invoice for the services (see Appendix I).

PART 4 – Project Management Structure

Kristen Erdem,
Director of
Business
Development,
Texas

Lauren
Kennedy,
Director of
Test
Development,
SLTI

City of
Austin

Claudia Diaz,
LC Manager,
Austin

Alexia Nguyen,
Manager,
Testing Center

Contract and project accountability:
Kristen Erdem, Director of Business Development, Berlitz Texas
Kristen will be the primary contact for all contract related items and will ultimately be accountable the entire project.

Test Customization and Consulting:
Lauren Kennedy, Director of Test Development, SLTI, a Berlitz Co.
Ms. Kennedy will be responsible for assessing the needs of the city of Austin at our kick-off meeting and will be responsible for recommending and developing any customization requirements for the City of Austin. Additionally, she will consult with the City of Austin on the establishment of cut scores (pass/fail scores).

Test Scheduling and Proctoring:
Claudia Diaz, Language Center Manager, Austin TX
Ms. Diaz will be the point of contact for scheduling/rescheduling of tests and for invoicing.

Test Rating and Results Delivery:
Alexia Nguyen, Manager, Testing Center,
Ms. Nguyen will be responsible for assigning raters and for supervising the rating and results delivery process.

PART 5 – Prior Experience

Berlitz and SLTI have a great deal of experience with the development and administration of language tests for students. In these projects, SLTI has played a number of roles, including principal test development contractor, item writing and item reviewing contractor, examiner, rator, and consultant. These varying responsibilities have allowed SLTI to demonstrate successful collaboration and a cooperative approach to client missions. An indicative sample of the projects the company has completed that are pertinent to the area of English and Spanish language proficiency testing and conducting test research is given below. For all test administration projects, Alexia Nguyen, Manager of the Berlitz Testing Center has been overseeing the the service to clients. Lauren Kennedy was Director of Test Development when all the test development projects were underway.

Translator Tests: City of Austin, Austin Energy (2012)

SLTI provided this utility company a customized translation test to qualify translators and translation companies that responded to an RFP. SLTI worked with the team at Austin Engergy to understand the tasks required of translators as well as their objectives for testing. We developed two forms of a translation test based specifically on the types of documents that translators will be expected to translate including texts ranging from the technical to those intended for a wide consumer audience. A scoring rubric was created with points awarded for the best translation that captured not only the proper translation, but also the proper register for the intended audience. We are making the test available at the Berlitz Language Center in Austin. The test will be scored at our testing center with support from SLTI.

Contact: Terry Nicholson, Senior Buyer
(512) 322-6586; terry.nicholson@austinenergy.com

Federal Court Interpreter Certification Exam: Administrative Office of the Courts (1996-Present)

In the years 1995 and 1996, SLTI was contracted to evaluate the Federal Court Interpreter Certification Exam (FCICE) for Spanish by the Administrative Office of the US Courts. After completing its evaluation, SLTI was contracted to develop an eight-year plan to improve the test program. In 2001, SLTI assumed responsibility for developing the written exam, which tests translation and interpretation skills in English and Spanish. SLTI developed new test specifications to make the exam more job-relevant, by simulating the types of tasks court interpreters actually do in the courtroom. Then, SLTI developed both the Spanish and English sections of the written exam using federally certified court interpreters to write the items. Since the initial round of item development, SLTI has continued to provide currency reviews for existing items, as well as develop and field test additional parallel forms of the test, with item writing occurring most recently in 2012.

Contact: Wanda Romberger, Manager, Court Interpreting Services
(757) 259-1879 or (800) 877-1233; wromberger@ncsc.dni.us

Medical Interpreter Tests: The California Endowment (2006-2011)

SLTI was contracted to design, implement, and coordinate the pilot testing of a series of Language Proficiency Tests and Interpreter Readiness Tests in Spanish, Cantonese, and Hmong for medical interpreters in California. The Spanish tests were originally developed to be implemented in a paper and pencil format. SLTI was contracted to evaluate these tests and made recommendations regarding their operationalization. SLTI has developed additional forms of the Spanish tests for implementation in a web -based delivery platform, including the development of all audio and video recordings. The web-based versions are currently being beta tested.

Contact: Ignatius Bau, Project Officer, The California Endowment
(415) 356-4342; IBau@Calendow.org

Berlitz Test of Speaking Skills: AT&T (2007 to present)

Berlitz conducts the Berlitz Test of Speaking Skills to AT&T. Each year we conduct thousands of Tests for call center agents around the world. Tests are conducted over the phone and results are delivered on the CEFR Scale. AT&T uses Berlitz to ensure that international representatives speak English clearly enough to communicate with AT&T's consumer customers in the United States.

Contact: Angela K. Richardson, Manager AT&T
(512) 465-4657, ar1324@att.com

Berlitz Test of Speaking Skills: CIGNA (2008 to present)

Berlitz provides the Berlitz Test of Speaking skills to Cigna insurance. We deliver this test over the phone to agents and employees of this organization. We provide the standard Berlitz Test of Speaking Skills on the CEFR scale plus additional customized components based upon roles for Customer Service and Medical terminology. Berlitz provides this test primarily for Spanish, but have also provided it for Tagalog and others. The purpose of testing is to ensure that representatives of CIGNA can appropriately provide service to non-English speaking clients of CIGNA.

Contact: Valencia Walker, Manager, CIGNA
valenciadenise@cigna.com

Berlitz Test of Speaking Skills/ Berlitz Test of Writing Skills: Scan Health Plan (2004 to present)

Berlitz provides Berlitz Test of Speaking Skills and Writing Skills to Scan. We deliver testing for: Spanish, Tagalog, Portuguese, Korean, Vietnamese, Mandarin, Japanese and others. We provide some customization for both medical and customer service according to their needs.

Berlitz Test of Speaking Skills and Customized Tests: Denver Public Schools (2011 to present)

Berlitz provides over 200 Spanish Tests per year for this system including the Berlitz Test of Speaking Skills and a proprietary writing test developed by the Denver Public Schools. This test is conducted to ensure the skills of Bi-lingual employee candidates with the Denver Public School system.

Contact: Dr. Nellie Cantu, Deputy Chief Academic Officer, Denver Public Schools 720-423-3932

PART 6 – Personnel

Kristen Erdem, Business Development Manager for Texas and Oklahoma
Kristen will be the primary contact for the City of Austin with respect to contract issues. Kristen joined Berlitz in 2011 as a Language Center Director and is now focused on sales and support of clients in her area. Kristen has a BA in French and Journalism from Louisiana State University. Prior to joining Berlitz, Kristen's experience included a wide array of client support roles in addition to teaching in the public schools. She is located at the Austin Language Center, 7000 North MoPac Expwy, Suite 200, Austin, TX 78733. She can be reached at 512-674-1957. E-mail: kristen.erdem@berlitz.us. The phone number for the Austin Langauge Center is 512-514-6238. (Kristen Erdem CV – See Appendix M)

Claudia Diaz, Language Center Manager for Austin, TX
Claudia will be responsible for scheduling and rescheduling of examinees as well as the proctoring of the tests at the Austin Languag Center. Claudia has been responsible for service at Berlitz in Austin since 2000. In that role she has been responsible for making and communicating the daily schedule at the Austin Language Center, She is responsible for ensuring the total satisfaction of Berlitz Customers. Claudia has a degree in Law from the University of Isdro Fabelo in Mexico. (Claudia Diaz CV – See Appendix M)

Alexia Nguyen, Berlitz Testing Center Manager
Alexia Nguyen is the Testing Center Manager. This is the primary facility for testing services throughout the country that oversees all Testing Agreements and scheduling of testing services. Alex's role will be to coordinate the scheduling of tests, the assignment of raters, and the delivery of results to the City of Austin. Alex has been in this position since 2003 and has worked directly with all of our testing clients over that time. Alex has an MA in English and has extensive experience working at Berlitz providing language services to clients. Before joining Berlitz in 1999, Alex was an interpreter and translator. Alex is Bilingual in English and French. She is based at our Testing Center in Woodland Hills CA at 6300 Canoga Ave. Suite 102, Woodland Hills, CA 91367, email: alexia.nguyen@berlitz.us. (Alexia Nguyen CV – See Appendix M)

Lauren Kennedy, Director of Test Development
Lauren Kennedy will act as the Program Manager (PM), overseeing the test customization and test evaluation. Lauren Kennedy is the Director of Test Development at SLTI. Since joining SLTI in August 2006, Ms. Kennedy has led teams of item writers who have developed reading passages and items for several tests of English skills to be used within professional and educational contexts. She also supervised the creation and implementation of a rater training program, and recently completed a study to evaluate its effectiveness. Prior to her work at SLTI, Ms. Kennedy worked on several test development projects including the Canadian Public Service Commission's Second Language Evaluation Test for Written Expression, the Canadian Academic English Language (CAEL) Assessment, and the Oral Proficiency Examination for Cooperative Education Students. She also collected focus-group data for a study investigating ESL students' Ontario Secondary School Literacy Test (OSSLT) test-taking processes. Ms. Kennedy earned her MA in Linguistics and Applied Language Studies and a Certificate in Teaching English as a Second Language from Carleton University in Ottawa. She has taught

English as a second language in academic and government settings. Additionally, she lived in Japan for one year as a high school exchange student and minored in Japanese as an undergraduate. She is currently pursuing a PhD in Linguistics and English Language (with a focus on Language Testing) at Lancaster University in the United Kingdom as a part-time, distance student. Ms. Kennedy is located at the SLTI headquarters at 6135 Executive Blvd., Rockville, MD 20852 and can be reached at 301-231-6046 ext. 1337, email: LKennedy@2lti.com. (Lauren Kennedy CV – See Appendix M)

Resumes are provided in Appendix M.

PART 7 – Proposal Acceptance Period
It is understood that all proposals are valid for a period of one hundred and twenty (120) calendar days subsequent to the RFP closing date unless a longer acceptance period is offered in the proposal.

PART 8 – Proprietary Information
It is understood all material submitted to the City becomes public property and is subject to the Texas Open Records Act upon receipt.

PART 9 – Authorized Negotiator
Kristen Erdem, the Berlitz Business Development Manager for Texas and Oklahoma, the authorized contact to negotiate contract terms and render binding decisions on contract matters. The Austin Language Center is at 7000 North MoPac Expwy, #200, Austin, TX, 78733. The phone number is 512-514-6238.

ATTACHMENT A: COMPLIANCE CHECKLIST

PROPOSERS ARE REQUESTED TO CONFIRM BELOW THAT PROPOSALS OFFERED THE SERVICES AT A MINIMUM TO THE SCOPE OF WORK REQUIREMENTS IN THE RFP AND THAT PROPOSAL PROVIDES DETAILS OF PROGRAM IN ACCORDANCE TO ALL REQUIREMENTS FOR THIS RFP PROJECT.

ITEM#	REQUIREMENT	**YES**	**NO**
1	Expanded services may be required with pricing for providing these optional services.	X	
2	Provide services for 36-month contract term, and with 1-36 month extension option periods	X	
3	re-examinations	X	
4	fixed price per test	X	
5	additional language requirements	X	
6	revision of test to address specialized skill sets used in the medical or legal work environment	X	
7	provide exam to an estimated 581 City of Austin employees	X	
8	Provide sample of auditory or visual recordable media	X	
9	Display sample of testing standards used for exam	X	
10	Sample process for scheduling exam	X	
11	Sample notification (pass/fail)	X	
12	Sample notification for rescheduling exam	X	
13.	Sample of guides for test takers	X	
14.	Sample reporting format	X	
15.	Confirmation of identification of fixed cite location	X	
16.	Review of test taking facility	X	
17.	A list of communication and other equipment	X	
18.	Sample invoice for services rendered	X	
19.	Sample of Spanish exam	X	
20.	Documents validating the degrees or teaching certificates of the exam administrators	X	
21.	Documents used to show the validity of the exam	X	
22.	Work hours of the exam administrators	X	
23.	After hour contact of the exam administrators	X	

I hereby attest that all of the information contained herein is accurate.

Cristin Edm 10/18/12

Signature Date

ATTACHMENT B: COST SUMMARY SHEET

Fees quoted shall be firm inclusive of all costs including but not limited to: labor, materials, supplies, printing services, copies, transportation costs prepaid and allowed to City destination, exam retesting services, administrative reports, supervision as necessary and administrative burden for providing the language proficiency exam services as required in the RFP.

Service Fees	Annual Est. Qty.	Exam Fee Per Employee	Year 1 Annual Cost	Year 2 Annual Cost	Year 3 Annual Cost
Spanish Language Proficiency Exam Services: Berlitz Test of Speaking Skills and Berlitz Test of Writing Skills with customization	150	$170	25,500	25,500	25,500
Non Spanish Language Proficiency Tests on list. No customization	Not established	$160	Dependant upon number of Examinees.		
Language Intiation Fee (For languages not on our list of current languages supported)	$350		Not an annual cost.		
Customization for each language outside of Spanish	$600 per Language		No annual cost.		

Kristen Erdem, Berlitz Languages Inc.
Corporate Name of Proposer

Kristen Erdem

Authorized Signature

10/18/12
Date

Appendix A: CEFR Language Proficiency Descriptors

A.1 SPEAKING

PROFICIENCY DESCRIPTORS: Overall Spoken Interaction, Grammatical Accuracy, General Linguistic Range, Phonological Control, Spoken Fluency
Source: *The Common European Framework of Reference for Languages, Council of Europe, 2002*

	[illegible]	[illegible]	[illegible]
Overall Spoken Interaction (holistic)	Has a good command of idiomatic expressions and colloquialisms with awareness of connotative levels of meaning. Can convey finer shades of meaning precisely by using, with reasonable accuracy, a wide range of modification devices. Can backtrack and restructure around a difficulty so smoothly the interlocutor is hardly aware of it.	Can express him/herself fluently and spontaneously, almost effortlessly. Has a good command of a broad lexical repertoire allowing gaps to be readily overcome with circumlocutions. There is little obvious searching for expressions or avoidance strategies; only a conceptually difficult subject can hinder a natural, smooth flow of language.	Can use the language fluently, accurately and effectively on a wide range of general, academic, vocational or leisure topics, marking clearly the relationships between ideas. Can communicate spontaneously with good grammatical control without much sign of having to restrict what he/she wants to say, adopting a level of formality appropriate to the circumstances. Can interact with a degree of fluency and spontaneity that makes regular interaction, and sustained relationships with native speakers quite possible without imposing strain on either party. Can highlight the personal significance of events and experiences, account for and sustain views clearly by providing relevant explanations and arguments.
Grammatical Accuracy (specific area)	Maintains consistent grammatical control of complex language, even while attention is otherwise engaged (e.g. in forward planning, in monitoring others' reactions).	Consistently maintains a high degree of grammatical accuracy; errors are rare and difficult to spot.	Good grammatical control; occasional 'slips' or non-systematic errors and minor flaws in sentence structure may still occur, but they are rare and can often be corrected in retrospect. Shows a relatively high degree of grammatical control. Does not make mistakes which lead to misunderstanding.
General Linguistic Range (specific area)	Can exploit a comprehensive and reliable mastery of a very wide range of language to formulate thoughts precisely, give emphasis, differentiate and eliminate ambiguity. No signs of having to restrict what he/she wants to say.	Can select an appropriate formulation from a broad range of language to express him/herself clearly, without having to restrict what he/she wants to say.	Can express him/herself clearly and without much sign of having to restrict what he/she wants to say. Has a sufficient range of language to be able to give clear descriptions, express viewpoints and develop arguments without much conspicuous searching for words, using some complex sentence forms to do so.
Phonological Control (specific area)	same as C1	Can vary intonation and place sentence stress correctly in order to express finer shades of meaning.	Has acquired a clear, natural, pronunciation and intonation.
Spoken Fluency (specific area)	Can express him/herself at length with a natural, effortless, unhesitating flow. Pauses only to reflect on precisely the right words to express his/her thoughts or to find an appropriate example or explanation.	Can express him/herself fluently and spontaneously, almost effortlessly. Only a conceptually difficult subject can hinder a natural, smooth flow of language.	Can communicate spontaneously, often showing remarkable fluency and ease of expression in even longer complex stretches of speech. Can produce stretches of language with a fairly even tempo; although he/she can be hesitant as he/she searches for patterns and expressions, there are few noticeably long pauses. Can interact with a degree of fluency and spontaneity that makes regular interaction with native speakers quite possible without imposing strain on either party.

Berlitz Testing

PROFICIENCY DESCRIPTORS: **Overall Spoken Interaction, Grammatical Accuracy, General Linguistic Range, Phonological Control, Spoken Fluency**
Source: *The Common European Framework of Reference for Languages, Council of Europe, 2002*

	B1	A2	A1
Overall Spoken Interaction (holistic)	Can communicate with some confidence on familiar routine and non-routine matters related to his/her interests and professional field. Can exchange, check and confirm information, deal with less routine situations and explain why something is a problem. Can express thoughts on more abstract, cultural topics such as films, books, music etc. Can exploit a wide range of simple language to deal with most situations likely to arise whilst traveling. Can enter unprepared into conversation on familiar topics, express personal opinions and exchange information on topics that are familiar, of personal interest or pertinent to everyday life (e.g. family, hobbies, work, travel and current events).	Can interact with reasonable ease in structured situations and short conversations, provided the other person helps if necessary. Can manage simple, routine exchanges without undue effort; can ask and answer questions and exchange ideas and information on familiar topics in predictable everyday situations. Can communicate in simple and routine tasks requiring a simple and direct exchange of information on familiar and routine matters to do with work and free time. Can handle very short social exchanges but is rarely able to understand enough to keep conversation going of his/her own accord.	Can interact in a simple way but communication is totally dependent on repetition at a slower rate of speech, rephrasing and repair. Can ask and answer simple questions. Initiate and respond to simple statements in areas of immediate need or on very familiar topics.
Grammatical Accuracy (specific area)	Communicates with reasonable accuracy in familiar contexts; generally good control though with noticeable mother tongue influence. Errors occur, but it is clear what he/she is trying to express. Uses reasonably accurately a repertoire of frequently used 'routines' and patterns associated with more predictable situations.	Uses some simple structures correctly, but still systematically makes basic mistakes – for example tends to mix up tenses and forget to mark agreement; nevertheless, it is usually clear what he/she is trying to say.	Shows only limited control of a few simple grammatical structures and sentence patterns in a learnt repertoire.
General Linguistic Range (specific area)	Has a sufficient range of language to describe unpredictable situations, explain the main points in an idea or problem with reasonable precision and express thoughts on abstract or cultural topics such as music and films. Has enough language to get by, with sufficient vocabulary to express him/herself with some hesitation and circumlocutions on topics such as family, hobbies and interests, work, travel, and current events, but lexical limitations cause repetition and even difficulty with formulation at times.	Has a repertoire of basic language which enables him/her to deal with everyday situations with predictable content, though he/she will generally have to compromise the message and search for words. Can produce brief everyday expressions in order to satisfy simple needs of a concrete type: personal details, daily routines, wants and needs, requests for information. Can use basic sentence patterns and communicate with memorized phrases, groups of a few words and formulae about themselves and other people, what they do, places, possessions etc. Has a limited repertoire of short memorized phrases covering predictable survival situations; frequent breakdowns and misunderstandings occur in non-routine situations.	Has a very basic range of simple expressions about personal details and needs of a concrete type.
Phonological Control (specific area)	Pronunciation is clearly intelligible even if a foreign accent is sometimes evident and occasional mispronunciations occur.	Pronunciation is generally clear enough to be understood despite a noticeable foreign accent, but conversational partners will need to ask for repetition from time to time.	Pronunciation of a very limited repertoire of learnt words and phrases can be understood with some effort by native speakers used to dealing with speakers of his/her language group.
Spoken Fluency (specific area)	Can express him/herself with relative ease. Despite some problems with formulation resulting in pauses and 'cul-de-sacs', he/she is able to keep going effectively without help. Can keep going comprehensibly, even though pausing for grammatical and lexical planning and repair is very evident, especially in longer stretches of free production.	Can make him/herself understood in short contributions, even though pauses, false starts and reformulation are very evident. Can construct phrases on familiar topics with sufficient ease to handle short exchanges, despite very noticeable hesitation and false starts.	Can manage very short, isolated, mainly pre-packaged utterances, with much pausing to search for expressions, to articulate less familiar words, and to repair communication.

Berlitz Testing

Appendix B: Sample Process for Scheduling Exam

Requests should be submitted to testing@berlitz.us

- Tests of **Spanish**

For a test to be scheduled and confirmed on a given day, it must be scheduled by no later 5 business days prior to the requested date.

- Tests of **French, German, Italian, Portuguese, Arabic and Thai**

For a test to be scheduled and confirmed on a given day, it must be scheduled at least 5 business days prior to the requested date.

Berlitz Testing confirms a schedule request by e-mail. In most cases, your requested date and time can be met. In some cases, the requested time cannot be met and an alternative time is offered. Please confirm acceptance or refusal of the alternative time.

Information you receive by e-mail in the schedule confirmation:

- Name of Examinee
- Test Language
- Contact information of Client Administrator
- Additional information your company needs for operations
- Location of Austin Berlitz Center

Appendix C: Sample Pass/Fail Score Report

Berlitz Testing

Score Report Berlitz Test of Speaking and Writing Skills

Name of Organization: City of Austin

Examinee Name: John Smith
Date of Test: October 1, 2012
Language: Spanish
Time of Test: 3:30pm
Test Location: Berlitz Center, Austin, TX

Berlitz Test of Speaking Skills

Pass	✓
Fail	

Berlitz Test of Writing Skills

Pass	✓
Fail	

Appendix D: Sample Notification for Rescheduling Exam

Rescheduling Tests

A test can be rescheduled without charge if Berlitz is notified by 5pm Central Time on the working day (Monday – Friday) prior to the scheduled date and time.

If the test is cancelled on the same day of the scheduled administration or if the Examinee is arrives more than 10 minutes after the scheduled time, the test will be noted as a "NO SHOW" and will be charged.

Appendix E: Sample Test-Taker Guide

Examinee Information & Instructions

Dear Examinee,
You are going to take a Berlitz Test. We look forward to you taking the test! The following questions should provide with the information you need.

Berlitz Test of Speaking Skills

What is the Berlitz Test of Speaking Skills? The Berlitz Test of Speaking Skills is a language test designed to measure your oral proficiency in a language.

How is it conducted? A Berlitz Testing "Rater", one of our language testers who has native proficiency in the test language, will lead a dialogue with you in the test language. This dialogue will last up to 20 minutes.

What is the topic of the dialogue in the test? The dialogue is based on a professional context. That means, the Rater will ask you questions about your experiences dealing with customers and co-workers or dealing with the workplace in general. The Rater's sole objective is to assess your language skills. Your rater will not test your knowledge about any profession or industry.

What should I do to prepare for the test? There is no specialized way to prepare for the test such as a specific course or study materials. You should simply be relaxed and prepared to listen and speak in the test language about your professional experiences.

What do I need to do to pass the test? The test is a pass/fail test with a cut score based upon the CEFR scale and established by the City of Austin. The test is designed to assess your ability to communicate orally in a professional context.

What if I don't understand a particular question or words on the test? Your Rater's objective is to listen to you speak. If you do not understand a question or words that s/he uses, your Rater will rephrase a question, use different words, or ask a new question. There is no penalty if you need to ask your Rater to repeat or rephrase any words or questions.

What if my Rater speaks a variety of the test language that I am not accustomed to (e.g., British vs. American English)? Berlitz Raters are trained to speak their variety of the test language properly, at a medium pace, and without a dialect or vocabulary that is not universally understood in the language. Our Raters neither reward nor penalize any use of the target language that is influenced by a specific variety of the test language.

What if I am very nervous at the start of the test? Don't worry! The Rater will begin the test with a few "warm up" questions about everyday topics so that you get used to speaking in the language.

What if there is a bad telephone connection and I cannot understand the Rater well? If there is static or any other sound on the telephone that prevents clear communication between you and the Rater, please tell the Rater and the s/he will terminate the call. Your test will be re-scheduled.

What do I need to do to set up the test? You do not have to do anything to set up the test. Your employer or prospective employer will arrange the test for you. You will be given a telephone in a quiet place so that you can take a call with your Rater at a specified date and time. Please be on time for the test. If you call your Dial-in more than 10 minutes after your scheduled time, you will be disqualified from taking the test.

Agreement to take the test
At the very beginning of the test the tester will read a disclaimer statement that will require your verbal agreement to participate in the test. The disclaimer asks for your permission to record the test. The rater will ask if you agree or disagree to being tested. You must agree to continue with the test.

Please Note: Your Berlitz Test of Speaking Skills will be audio-recorded and kept on file with BerlitzTesting. The recording may be used exclusively for quality assurance or review by your (prospective) employer.

Berlitz Test of Writing Skills

How long does the Berlitz Test of Writing Skills take?
The Writing Skills Test takes no more than 60 minutes.

Where will I take the test?
You will take the test in a room assigned for this purpose in your organization.

What should I take with me?
Please take a ball point pen.

What is the content of the test?
Typical tasks for a test of writing skills include:

- writing a response in the target language to an email also written in the target language which is consistent to its tone and format
- writing a short essay in the target language about a non-technical topic set in a general business / service-oriented context

If the test includes, reading skills too, the following are typical tasks:

- reading a passage in the target language and responding orally to questions about it
- reading a passage in the target language and answering multiple choice questions about it

Appendix G: Review of Test-Taking Facility

The test taking facility is a Berlitz Center located in Austin. Berlitz has maintained a presence in the city of Austin for more than ten years. Our facilities are quiet, clean, professional environments well suited to quiet study or assessment. The Center is staffed full-time by Berlitz Employees whose sole purpose is to provide superior customer service and a professional environment.

Appendix H: A list of communication and other equipment

Equipment needed for Test of Speaking Skills:
Telephone
Digital recording devices for face-to-face examinations
Computers for uploading the audio recordings

Equipment needed for Test of Writing Skills:
None

Appendix I: Sample Invoice

Invoice 1

Date: September 28, 2012

Invoice to: The City of Austin
XXXX
Email: XXXX

From: Berlitz Language Center: Austin

Reference: Language Testing

Description of Services: Proficiency Testing of Employees

Examinee	Date of Exam	Language Tested	Location	Price
Name				

Total Amount Due **XXXX**

REMIT TO: **Berlitz Language Center**

7000 North MoPac Suite 200
Austin, TX 78731

Appendix J: Sample Requirements for Exam Administrators and Sample Exam Administrator Resume

Berlitz Raters (Evaluators):

Language Proficiency: applicants must have native language proficiency in the language or languages they administer tests for. By 'native language proficiency' we mean the ability to communicate in a language as a mother-tongue with no noticeable influence of another language. Also, successful applicants will be accustomed to speaking the language effectively in professional situations.

Language Teaching/Testing Experience: Applicants have some combination of skills including language teaching or language testing.

Educational Qualifications: Applicants must hold at least a baccalaureate degree (4-year college degree).

Training: We invite successful applicants to our initial Rater Training Course that is mandatory. The applicants, as trainees, must successfully complete the training before they may begin as Raters for Berlitz. As part of the training, Raters must pass a test showing their calibration to the CEFR.

OBJECTIVE

To teach Spanish to all that want to speak this language for travel, pleasure or for future career growth.

EDUCATION AND AWARDS

Colegio del los Sacrados Corazones; Quito, Ecuador, South America G.P.A. 4.000
Berlitz School of Languages: Austin, Texas
Berlitz Teacher of the Year: Awarded five (5) times over twenty-year period.
Maximillian Berlitz Award. Awarded in 2008 for distinguished professionalism and dedication.

Relevant Courses

Native-born speaker in Spanish with excellent grades in Spanish grammar, literature and history.
Berlitz intensive training on the Berlitz Method in 1988, with continuous training on the latest language teaching techniques and materials over the past twenty (20) years.

WORK EXPERIENCE

1. Twenty years teaching the Berlitz Method to over 5,000 students of different walks of life, professions, skills and interests under the following programs:
   Total Emergence
   Large Groups
   Medium Groups
   Small Groups
   Children Groups
   Individual
2. Berlitz Group Coordinator for six (6) years for all languages taught by Berlitz.
3. Simultaneous translator at medical conferences and legal depositions.
4. Evaluations via telephone of applicants who need to demonstrate their command of the Spanish language as part of their employment.

Appendix K: Testing Center Office Hours

The office hours for the Testing Centers are Monday through Friday, 8am to 5pm. These hours can be extended upon request and with prior scheduling.

Appendix L: After Hours Contact Email

For requests and issues after hours, Berlitz provides an e-mail that is monitored: afterhours@berlitz.us

Appendix M: Curricula Vitae

Lauren Kennedy
6135 Executive Blvd.
Rockville, MD 20852, USA
LKennedy@2LTI.com
Office: 301-231-6046 x1337

Education

Ph.D. Candidate Linguistics and English Language (2010-Present)
Lancaster University *(Lancaster, UK)*
Focus on language testing and assessment
Part-time by distance
Working thesis title: *Exploring anchor-based methods for judgementally estimating item difficulty in English for academic purposes reading test items*

M.A. Applied Language Studies (2007)
Carleton University *(Ottawa, Ontario)*
Focus on language testing
M.A. research paper:
Kennedy, L. (2007). *Expanding test specifications with rhetorical genre studies and Activity Theory analyses*. (ERIC Document Reproduction Service No. ED498923).

North American Mobility Program Graduate Exchange (2005)
Portland State University *(Portland, Oregon)*
Focus on language testing and research methods

Certificate in Teaching English as a Second Language (2003)
Carleton University *(Ottawa, Ontario)*

Bachelor of Arts and Social Science Honours in Linguistics and Applied Language Studies (2003)
Carleton University *(Ottawa, Ontario)*
Minor in Business
Minor in Japanese Language

Additional training

American Council on the Teaching of Foreign Languages (ACTFL) Oral Proficiency Interview (OPI) Assessment Workshop (2010)
ACTFL *(Boston, Massachusetts)*

3-day workshop on ILR Text Typology and Passage Rating (receptive skills) (2008 and 2010)
2-day orientation to DLPT 5 Development and Review (2008 and 2010)
Defense Language Institute *(Monterey, California)*

Introduction to Business Statistics (2007)
Montgomery College *(Rockville, Maryland)*

Selected language testing and teaching experience

Director, Test Development (2006-Present)
Second Language Testing, Inc., Acquired by Berlitz Languages in 2011 *(Rockville, Maryland)*
Technical Expertise

- Directed the recruitment, training, test development, and item review of language tests in over 15 languages for the US Department of Defense (Defense Language Proficiency Tests 5th Edition (DLPT 5), Very Low Range (VLR) tests, and Translation and Interpretation Tests).
- Lead the review of the US Department of State FSI Reading Proficiency Test and recommended revisions that were subsequently implemented by the FSI.
- Managed test development and item review projects (see *Testing Systems Experience Below*).
- Designed a standardized test development process to ensure that all test and scoring materials were fair, reliable, and valid, and in compliance with industry standards and client requirements.
- Worked closely with subject matter experts to design, develop, and review test specifications, test content, scoring rubrics and procedures.
- Recruited, trained, supervised, and evaluated item writers for adherence to the requirements of the test specifications, work order, and test development procedures.
- Developed and reviewed proficiency, aptitude, and achievement test items and materials.
- Created support materials and instructional documents for clients, test takers, and score users.
- Designed and reviewed the content and format of score reports to ensure their interpretations are appropriate, useful, and accessible to score users.
- Preformed item analysis for item review, item banking, and test assembly purposes.
- Facilitated standard setting and cut score studies.
- Preformed and evaluated test form equating.
- Conducted validation research to evaluate the appropriateness and usefulness of test score interpretations and decisions.
- Wrote technical and research reports.
- Researched and prepared white papers on topics relevant to assessment.
- Instructed and developed staff to ensure their understanding of test and measurement issues so that products produced were of the highest quality.
- Advised test review committee members and item writers on assessment issues.
- Presented public webinars to teachers about test development.
- Monitored contemporary issues, trends, and research in language testing and assessment in order to develop items and tests based on client needs and strategic business needs.
- Represent SLTI and Berlitz Languages at professional meetings and conferences.

Impact of the Ontario Secondary School Literacy Test on Second Language Students (2006)
Research Assistant to Dr. Janna Fox *(Language Assessment and Testing Research Unit, Carleton University, Ottawa, Ontario)*

- Designed focus group questions
- Moderated focus groups
- Transcribed focus group data

Second Language Evaluation – Written Expression Item Writer (2006)
Government of Canada, Public Service Commission (*Ottawa, Ontario)*

- Wrote items for conformity with specifications and prototypes as well as good item writing practices

Test Item Writer, Researcher, Administrator, and Rater (2003-2006)

Canadian Academic English Language (CAEL) Assessment *(Carleton University, Ottawa, Ontario)*

- Wrote and revised the Canadian Academic English Language (CAEL) Assessment Test Score and Users' Guide
- Designed and researched tests as part of a test development team
- Prepared test materials
- Administered and rated tests

Intensive ESL Teacher (2005)
Portland State University *(Portland, Oregon)*

- Taught Reading and Listening-Speaking courses

Intensive ESL Teacher (2003-2005)
Carleton University *(Ottawa, Ontario)*

- Taught Core (integrated skills), Writing Process, Test-taking Skills, and Business Communication courses

ESL Teacher (2004)

- Government of Canada, Canadian International Development Agency *(Gatineau, Quebec)*

Test Item Writer, Researcher, and Rater (2003-2004)
Oral Proficiency Examination for Cooperative Education Students (*Carleton University, Ottawa, Ontario*)
Developed test specifications and items
Revised previous test versions
Prepared test materials
Evaluated and analyzed examinee performance

Refereed publications

Ackermann, K., L. Kennedy. (2010). Research notes: Standardizing rater performance: Empirical support for regulating language proficiency test scoring. *Pearson Research Notes, September 2010*, http://pearsonpte.com/research/Documents/Research_Notes_Standardizing_Rater_Perfomance_1Sept10_KA.pdf

Kennedy, L. and C. Stansfield. (2010). The Reading Proficiency Interview (RPI): A rapid response test development model for assessing reading proficiency on the ILR Scale. *Applied Language Learning, 20* (1&2), 1-16. http://www.dliflc.edu/file.ashx?path=archive/documents/ALL20c.pdf

Kennedy, L. (2010). Universally Designed content assessments for English learners. *AccELLerate!, 3* (1), 12-13. http://www.ncela.gwu.edu/files/uploads/17/Accellerate_3_1.pdf

Wang, J. (Ed.). (2010). The official guide to the Pearson Test of English Academic. Hong Kong: Pearson Longman Asia ELT. Lead contributor to Chapters 3-6.

Kennedy, L. (2006). ESP: English for Scouting purposes. *The Canadian Leader Magazine June/July*, 8-9.

Refereed presentations

Ackermann, K., L. Kennedy, J. de Jong, and Y. Zheng. (2010). *Standardizing rater performance: Empirical support for regulating language proficiency test scoring.* Paper presented at the 7th Conference of the International Test Commission, July 19-21, 2010.

Kennedy, L., J. de Jong, C. Duron, and C. Casteel. (2008). *Improving a rater training program: Lessons learned from the Pearson Test of English Academic field tests.* Paper presented at the East Coast Organization of Language Testers Conference, November 7-8, 2008.

Kennedy, L. and C. Stansfield. (2009). *The Reading Proficiency Interview (RPI): A rapid responses test development model for assessing reading proficiency on the ILR Scale.* Poster presented at the East Coast Organization of Language Testers Conference, November 6-7, 2009.

Kennedy, L. (2007). *Expanding test specifications with rhetorical genre studies and activity theory analyses.* Paper presented at The Reading Matrix Online Conference http://www.readingmatrix.com, March 2, 2007.

Invited presentations and workshops

Kennedy, L. and C. Casteel. (2010). *National Clearinghouse for English Language Acquisition Webinar: Test Development and Item Writing for Teachers of ELLs.* Online Webinar http://www.ncela.gwu.edu/webinars/event/18/, March 31, 2010.

Stansfield, C., L. Kennedy, and K. Liu. (2010). *National Clearinghouse for English Language Acquisition Webinar: Using Universal Design in Test Development to Benefit ELLs.* Online Webinar http://www.ncela.gwu.edu/webinars/event/24/, June 10, 2010.

Technical reports

Stansfield, C., A. Evers, M. Taha, J. Turner, H. Pereira, J. Gao, L. Kennedy. (2010) *English to Arabic Translator/Interpreter Screening Test & Arabic to English Translator/Interpreter Screening Test.* Technical Report to Global Linguist Solutions: Second Language Testing, Inc.

Kennedy, L., J. Miles, and C. Stansfield. (2008). *Evaluation of the Foreign Service Institute Reading Proficiency Test.* Technical report to the United States Department of State, Foreign Service Institute: Second Language Testing, Inc.

Kennedy, L., H. Farhady, C. Duron, C. Casteel, A. Dudley, A. Powers, M. Fruit, and C. Stansfield. (2008). *Pearson Academic English Test supervisor and marker training evaluation report: First AET field test (Fall 2007).* Technical report to Pearson Language Assessments: Second Language Testing, Inc.

Kennedy, L. and L. Hart-Gonzalez. (2007). *Partially correct distractors in multiple-choice items: What the literature says.* Technical report to the Defense Language Institute Foreign Language Center (DLIFLC): Second Language Testing, Inc.

Kennedy, L., C. Stansfield, and J. Kelly. (2007). *Universal design and plain language in NAEP: Theory and practice.* Technical report to NAEP Education Statistics Service Institute and the National Center for Educational Statistics: Second Language Testing, Inc.

Stansfield, C., J. Kelly, L. Kennedy, D. Kenyon, and M. Louguit. (2007). *Test specifications for the Defense Language Aptitude Battery*. Technical report to the Center for the Advanced Study of Language (CASL) and the Defense Language Institute Foreign Language Center (DLIFLC): Second Language Testing, Inc.

Kennedy, L. (2006). *A review of literature on response choice writing guidelines for multiple-choice items*. Technical report to NAEP Education Statistics Service Institute and the National Center for Educational Statistics: Second Language Testing, Inc.

Testing systems experience

ACTFL Oral Proficiency Interview (OPI)
Benesse GTEC for STUDENTS
Berlitz Test of Listening and Reading (BTLR)
Berlitz Test of Speaking Skills
Berlitz Test of Writing Skills
Canadian Academic English Language (CAEL) Assessment
Defense Language Aptitude Battery (DLAB)
Defense Language Proficiency Test 5th Edition (DLPT5) – Multiple Choice and Constructed Response
English to Arabic Translator/Interpreter Screening Test & Arabic to English Translator/Interpreter Screening Test
Federal Bureau of Investigation (FBI) Listening Summary Translation Exam
Government of Canada Second Language Evaluation – Written Expression (English)
LAS-Links
National Assessment of Educational Progress (NAEP)
National Geographic Reading Edge Assessment
New Mexico Standards Based Assessment: Spanish version
Pearson Test of English Academic
Pearson Test of English General
Reading Proficiency Interview (RPI)
Simulated Oral Proficiency Interview (iSPEAK)
Tests of language aptitude (MLAT, CB-MLAT, PLAB, MLAT-E)
United States Federal Court Interpreter Certification Exam (FCICE)
United States Foreign Service Institute (FSI) Reading Proficiency Test
Very Low Range (VLR) Defense Language Proficiency Test

Professional service

Reviewer, *Language Testing*
Reviewer, *East Coast Organization of Language Testers 2012 Conference (ECOLT)*
Member of ASTM International Subcommittee F15.64 Language Proficiency
Interagency Language Roundtable (ILR) Listening Summit delegate (2009)
Interagency Language Roundtable (ILR) Oral Proficiency Interview (OPI) Summit delegate (2008)
Student organizer of Language Testing Research Colloquium (LTRC) 2005

Professional memberships

American Council on the Teaching of Foreign Languages (ACTFL)
ASTM International
International Language Testing Association (ILTA)
East Coast Organization of Language Testers (ECOLT)
Interagency Language Roundtable (ILR)

Berlitz Testing

KRISTEN PEDERSEN ERDEM

512-674-1957 kperdem@yahoo.com

PROFESSIONAL SUMMARY

Achieved excellence in sales, sales management, and sales support positions. Experienced in technical sales for the wireless industry. Practiced in the art of consultative sales process. Proven track record in sales and marketing in various American markets, often exceeding sales goals by as much as 400% of quota. Fluent in French (read, write, speak).

EMPLOYMENT HISTORY

Business Development Manager

9/2011 - present

Berlitz Inc., Austin, TX

Consultative sales for the State of Texas, Oklahoma, & Arkansas for clients with a global reach. Responsible for entire sales cycle including prospecting, consultative process, proposal development, negotiation, closing, and initial implementation for language and cultural training programs and services.

- Sold upwards of $250k in language assessment and testing services for local government entities
- Sold $200k in language training services for a global e-commerce client
- 58% growth in target market from 50 accounts to 85 accounts in 6 months

Account Executive 2010 - 2011

Community Coffee Company LLC, Austin, TX

Direct B2B sales in central Texas for the Food and Beverage Market

- Achieved 400% of quota in 1st quarter of employment
- Expanded territory into Office Coffee Service from 0 to 25 accounts increasing footprint beyond Restaurant and Food Service clients
- Ranked consistently a monthly basis in top 3 for Sales of Ancillary Products, which represent profit margins of approximately 30%

Account Executive 2004 - 2010

Sprint Nextel, New Orleans, LA & Austin, TX

Corporate sales and account management as a trusted advisor to organizations in a variety of industries.

- Achieved 400% growth in target market from 5 to 91 clients the first year in Austin by cultivating relationships within small and medium sized businesses
- Recognized as Top Solution Seller 12 times within a five-year period with as much as 200% of data quota achievement
- Awarded Top Account Executive during a three-year period for General Business Accounts for outstanding sales performance

Elementary Teacher 2003 - 2004

Jefferson Parish Public Schools, Kenner, LA

Taught 3rd grade to 23 students in a public school in an impoverished suburb of New Orleans

- Met all parents within the first month of school and spoke to every parent or guardian on a rotating three week period
- Team-taught with Special Ed teacher to provide an inclusive teaching environment

- Advocated for late arrival students to be allowed to receive their federally funded breakfast in the classroom

Customer Service Manager & Account Manager 1993 - 2002

AirTouch Paging/Bell Atlantic Messaging Services/Verizon Wireless, San Francisco, CA & Syracuse, NY

Promoted from Customer Service Manager to Sales Account Manager, and then transferred to East Coast.

- Led by example my service team of 12 individuals, exceeding sales quota every month
- Developed training manual for service team that was adopted by the Western region
- Took on a Strategic sales role of "hunting" for new business in previously untapped markets of the San Francisco Financial District and the market territory of Syracuse, New York, consistently exceeding sales revenue goals by 50 to 150%

EDUCATION & SALES TRAINING

Louisiana State University, Baton-Rouge, LA

Bachelor of Arts, French and Journalism 1991

- Rotary Scholar Post-Graduate Year Abroad. Montpellier, France (1992 – 1993)
- Study Abroad, L'Universite de l'Etat, Mons, Belgium (1989 – 1990)
- Sprint University covering modules in technology and consultative selling method (2004-2010)

TECHNICAL SKILLS

CRM Systems: Salesforce.com, GoldMine, Act!

Wireless Technology: Blackberry, Android, Windows Mobile, 3G/4G, Cisco

Productivity Software: Microsoft Office 2007 (Word, Excel, PowerPoint), Outlook, Skype

Business Information Tools: Hoover's, ResourceUSA, Vault, GlassDoor, Google Alerts

Social Media: LinkedIn, Facebook, Twitter

COMMUNITY SERVICE

Keep Austin Gluten Free to benefit Alamo Celiac, 2010 & 2011

- Launched and organized vendor fair

St John Neumann Catholic Church, 2009 - present

Tops Soccer Program, Westlake Youth Soccer, 2008 - 2009

ESL teacher, Bati Koleji, Istanbul, Turkey, 2002 – 2003

Alexia Nguyen
Email: *alexia.nguyen@berlitz.us*

EDUCATION

May 1998	*Master of Arts, English. University of Arts and Science, Nice, France.*
Dec. 1992	*Certificate of Proficiency in English. Chichester, UK.*

PROFESSIONAL EXPERIENCE

2003-present	*Manager of Berlitz Testing Center, Woodland Hills, CA. Duties include managing operations for a Language Testing Center, customer service to current clients and sales support for potential clients.*
Oct 1999-2003	*Service/Accounting Associate, Payroll Manager at the Beverly Hills and Santa Monica Berlitz Language Centers. Responsibilities include customer service and collection/billing for the centers' corporate accounts as well as payroll for all staff members (administrative and teaching).*
1999-2003	*Translation (French-English, English-French) of official documents such as contracts and business correspondence for various companies and individuals including executives of Smart&Final and Casino USA.*
Summers & winters 1992-1998	*Interpreter and translator for various shows (MIP TV, FIF) at the Cannes Palais des Festivals (France). Included helping exhibitors communicate with potential buyers/investors and translating various documents.*
1993-1997	*Interpreter and translator for various seminars and conventions at the Nice Palais des Congrès (France): welcoming foreign guests, touring the congress in English, interpreting lectures and meetings and translating documents and pamphlets.*

LANGUAGES

Fluent in French and English (oral & written).
Some knowledge of Spanish and Italian (oral & written).

Claudia Diaz
11500 Jollyville Rd. #714
Austin, TX 78759
Cell: 512-913-6425
Claudia.diaz@berlitz.us

Education

2011- Present **Paralegal Coursework**
Austin Community College Austin, TX

1985-1990 **Bachelor Degree**
Attorney at Law Mexico, Mex.
Universidad Isdro Fabela

Experience

2000 – Present **Language Center Manager**
Berlitz Language Center Austin, TX
Language Center Operation, Administrative Functions, Costumer Service.

2001 – Present **Merchandise Coordinator/Key Carrier**
Tjmaxx Austin, TX
Conduct merchandising plans and priorities. Markdowns, recovery and maintenance of store.

1992 – 1997 **Civil Lawyer**
Civil Office Mexico, Mex
Employment Law, Family Law

Languages
Spanish (native Language)

Skills
Microsoft Word, Excel, and Power Point
Payroll
Scheduling employees hours

Appendix N: List of Language Tests Offered (Speaking and Writing)

Speaking Skills	Writing Skills
Arabic	Arabic
Assyrian	Dutch
Armenian	English
Bosnian	Farsi
Bulgarian	French
Cantonese	German
Cambodian	Italian
Croatian	Portuguese
Dutch	Spanish
English	Tagalog
French	Thai
German	
Greek	
Hebrew	
Hindi	
Italian	
Japanese	
Korean	
Malay	
Mandarin	
Moldavian	
Norwegian	
Polish	
Portuguese	
Romanian	
Russian	
Serbian	
Slovak	
Spanish	
Tagalog	
Thai	
Vietnamese	

Appendix O: Certifications

O.1 Terms and Conditions
O.2 Purchasing Office Reference Sheets
O.3 Equal Employment/ Fair Housing Office Non-Discrimination Certification
O.4 Non-Suspension or Debarment Certification
O.5 City of Austin Non-Collusion, Non-Conflict of Interest, and Anti-Lobbying Affidavit
O.6 City of Austin, Texas Living Wages and Benefits Contractor Certifications
O.7 City of Austin, Nonresident Bidder Provisions
O.8 City of Austin Local Business Presence Identification Form
O.9 Minority- and Women-Owned Business Enterprise Procurement Program No Goals Form

O.1 Terms and Conditions

Berlitz complies with all Federal, State, and Local laws and regulations, and this proposal complies with all terms and conditions of Request for Proposal No. JSD0129.

Name: KRISTEN ERDEN

Signature: Kristen Erden

Date: 10/18/12

O.2 Purchasing Office Reference Sheets

CITY OF AUSTIN
PURCHASING OFFICE
REFERENCE SHEET
Please Complete and Return This Form with the Offer

Solicitation Number: JSD0129

Offeror's Name Berlitz Languages, Inc. Date 10/11/12

The Offeror shall furnish, with the Offer, the following information, for at least 5 recent customers to whom products and/or services have been provided that are similar to those required by this Solicitation.

To add additional references to this form, click the Add Reference Button. ========> Add Reference

Company's Name	AT&T		
Name of Contact	Angela Richardson	Contact Title	Manager
Present Address			
City		State	Zip Code
Telephone Number	(512) 465-4657	FAX Number	(512) 870-4749
Email Address	ar1324@att.com		

CITY OF AUSTIN
PURCHASING OFFICE
REFERENCE SHEET
Please Complete and Return This Form with the Offer

Solicitation Number: JSD0129

Offeror's Name: Berlitz Languages, Inc. Date: 10/11/12

The Offeror shall furnish, with the Offer, the following information, for at least 5 recent customers to whom products and/or services have been provided that are similar to those required by this Solicitation.

To add additional references to this form, click the Add Reference Button. ========> Add Reference

Company's Name	CIGNA		
Name of Contact	Valencia Walker	Contact Title	
Present Address			
City		State	Zip Code
Telephone Number		FAX Number	
Email Address	valenciadenise@cigna.com		

CITY OF AUSTIN
PURCHASING OFFICE
REFERENCE SHEET
Please Complete and Return This Form with the Offer

Solicitation Number: JSD0129

Offeror's Name: Berlitz Languages, Inc. Date: 10/11/12

The Offeror shall furnish, with the Offer, the following information, for at least 5 recent customers to whom products and/or services have been provided that are similar to those required by this Solicitation.

To add additional references to this form, click the Add Reference Button. ========> Add Reference

Company's Name	Denver Public Schools		
Name of Contact	Dr. Nellie Cantu	Contact Title	Deputy Chief Academic Officer
Present Address	Suite 600, 900 Grant Street		
City	Denver	State	CO Zip Code 80203
Telephone Number	(720) 423-3932	FAX Number	
Email Address			

Berlitz Testing

CITY OF AUSTIN
PURCHASING OFFICE
REFERENCE SHEET
Please Complete and Return This Form with the Offer

Solicitation Number: JSD0129

Offeror's Name: Berlitz Languages, Inc. Date: 10/11/12

The Offeror shall furnish, with the Offer, the following information, for at least 5 recent customers to whom products and/or services have been provided that are similar to those required by this Solicitation.

To add additional references to this form, click the Add Reference Button. ========> Add Reference

Company's Name	City of Austin, Austin Energy		
Name of Contact	Terry Nicholson	Contact Title	Senior Buyer
Present Address			
City		State	Zip Code
Telephone Number	(512) 322-6586	FAX Number	(512) 322-6490
Email Address	terry.nicholson@austinenergy.com		

CITY OF AUSTIN
PURCHASING OFFICE
REFERENCE SHEET
Please Complete and Return This Form with the Offer

Solicitation Number: JSD0129

Offeror's Name: Berlitz Languages, Inc. Date: 10/11/12

The Offeror shall furnish, with the Offer, the following information, for at least 5 recent customers to whom products and/or services have been provided that are similar to those required by this Solicitation.

To add additional references to this form, click the Add Reference Button. =======> Add Reference

Company's Name	National Center for State Courts		
Name of Contact	Wanda Romberger	Contact Title	Manager, Court Interpreting Services
Present Address			
City		State	Zip Code
Telephone Number	(800) 877-1233	FAX Number	
Email Address	wromberger@ncsc.dni.us		

O.3 Equal Employment/ Fair Housing Office Non-Discrimination Certification

City of Austin, Texas
EQUAL EMPLOYMENT/FAIR HOUSING OFFICE
NON-DISCRIMINATION CERTIFICATION

SOLICITATION NO JSD0129

City of Austin, Texas
Human Rights Commission

To: City of Austin, Texas, ÒOWNERÓ:

I hereby certify that our firm conforms to the Code of the City of Austin, Section 5-4-2 as reiterated below:

Chapter 5-4. Discrimination in Employment by City Contractors.

Sec. 4-2 Discriminatory Employment Practices Prohibited. As an Equal Employment Opportunity (EEO) employer, the Contractor will conduct its personnel activities in accordance with established federal, state and local EEO laws and regulations and agrees:

(B) (1) Not to engage in any discriminatory employment practice defined in this chapter.

(2) To take affirmative action to ensure that applicants are employed, and that employees are treated during employment, without discrimination being practiced against them as defined in this chapter. Such affirmative action shall include, but not be limited to: all aspects of employment, including hiring, placement, upgrading, transfer, demotion, recruitment, recruitment advertising; selection for training and apprenticeship, rates of pay or other form of compensation, and layoff or termination.

(3) To post in conspicuous places, available to employees and applicants for employment, notices to be provided by OWNER setting forth the provisions of this chapter.

(4) To state in all solicitations or advertisements for employees placed by or on behalf of the Contractor, that all qualified applicants will receive consideration for employment without regard to race, creed, color, religion, national origin, sexual orientation, gender identity, disability, veteran status, sex or age.

(5) To obtain a written statement from any labor union or labor organization furnishing labor or service to Contractors in which said union or organization has agreed not to engage in any discriminatory employment practices as defined in this chapter and to take affirmative action to implement policies and provisions of this chapter.

(6) To cooperate fully with OWNERÕs Human Rights Commission in connection with any investigation or conciliation effort of said Human Rights Commission to ensure that the purpose of the provisions against discriminatory employment practices are being carried out.

(7) To require compliance with provisions of this chapter by all subcontractors having fifteen or more employees who hold any subcontract providing for the expenditure of $2,000 or more in connection with any contract with OWNER subject to the terms of this chapter.

For the purposes of this Offer and any resulting Contract, Contractor adopts the provisions of the CityÕs Minimum Standard Nondiscrimination Policy set forth below.

City of Austin
Minimum Standard Non-Discrimination in Employment Policy:

As an Equal Employment Opportunity (EEO) employer, the Contractor will conduct its personnel activities in accordance with established federal, state and local EEO laws and regulations.

The Contractor will not discriminate against any applicant or employee based on race, creed, color, national origin, sex, age, religion, veteran status, gender identity, disability, or sexual orientation. This policy covers all aspects of employment, including hiring, placement, upgrading, transfer, demotion, recruitment, recruitment advertising, selection for training and apprenticeship, rates of pay or other forms of compensation, and layoff or termination.

Further, employees who experience discrimination, sexual harassment, or another form of harassment should immediately report it to their supervisor. If this is not a suitable avenue for addressing their complaint, employees are advised to contact another member of management or their human resources representative. No employee shall be discriminated against, harassed, intimidated, nor suffer any reprisal as a result of reporting a violation of this policy. Furthermore, any employee, supervisor, or manager who becomes aware of any such discrimination or harassment should immediately report it to executive management or the human resources office to ensure that such conduct does not continue.

Contractor agrees that to the extent of any inconsistency, omission, or conflict with its current non-discrimination employment policy, the Contractor has expressly adopted the provisions of the CityÕs Minimum Non-Discrimination Policy contained in Section 5-4-2 of the City Code and set forth above, as the ContractorÕs Non-Discrimination Policy or as an amendment to such Policy and such provisions are intended to not only supplement the ContractorÕs policy, but will also supersede the ContractorÕs policy to the extent of any conflict.

UPON CONTRACT AWARD, THE CONTRACTOR SHALL PROVIDE A COPY TO THE CITY OF THE CONTRACTORÕS NON-DISCRIMINATION POLICY ON COMPANY LETTERHEAD, WHICH CONFORMS IN FORM, SCOPE, AND CONTENT TO THE CITYÕS MINIMUM NON-DISCRIMINATION POLICY, AS SET FORTH HEREINOR **THIS NON-DISCRIMINATION POLICY, WHICH HAS BEEN ADOPTED BY THE CONTRACTOR FOR ALL PURPOSES (THE FORM OF WHICH HAS BEEN APPROVED BY THE CITYÕS EQUAL EMPLOYMENT/FAIR HOUSING OFFICE), WILL BE CONSIDERED THE CONTRACTORÕS NON-DISCRIMINATION POLICY WITHOUT THE REQUIREMENT OF A SEPARATE SUBMITTAL.**

Sanctions:
Our firm understands that non-compliance with Chapter 5-4 may result in sanctions, including termination of the contract and suspension or debarment from participation in future City contracts until deemed compliant with the requirements of Chapter 5-4.

Term:
The Contractor agrees that this Section 0800 Non-Discrimination Certificate or the ContractorÕs separate conforming policy, which the Contractor has executed and filed with the Owner, will remain in force and effect for one year from the date of filing. The Contractor further agrees that, in consideration of the receipt of continued Contract payments, the ContractorÕs Non-Discrimination Policy will automatically renew from year-to-year for the term of the underlying Contract.

Dated this 11th day of October, 2012.

CONTRACTOR Kristen Erdem, Berlitz Languages, Inc.

Authorized Signature Kristen Erdem

Title Business Development Manager for Texas and Oklahoma

O.4 Non-Suspension or Debarment Certification

City of Austin, Texas
NON-SUSPENSION OR DEBARMENT CERTIFICATION

SOLICITATION NO. USD0129

The City of Austin is prohibited from contracting with or making prime or sub-awards to parties that are suspended or debarred or whose principals are suspended or debarred from Federal, State, or City of Austin Contracts. Covered transactions include procurement contracts for goods or services equal to or in excess of $25,000.00 and all non-procurement transactions. This certification is required for all Vendors on all City of Austin Contracts to be awarded and all contract extensions with values equal to or in excess of $25,000.00 or more and all non-procurement transactions.

The Offeror hereby certifies that its firm and its principals are not currently suspended or debarred from bidding on any Federal, State, or City of Austin Contracts.

Contractor's Name:	Berlitz Languages, Inc.		
Signature of Officer or Authorized Representative:	Kristen Erdem	Date:	10/11/12
Printed Name:	Kristen Erdem		
Title:	Business Development Manager for Texas and Oklahoma		

O.5 City of Austin Non-Collusion, Non-Conflict of Interest, and Anti-Lobbying Affidavit

CITY OF AUSTIN
NON-COLLUSION,
NON-CONFLICT OF INTEREST, AND ANTI-LOBBYING AFFIDAVIT
SOLICITATION NO. USD0129
FOR

Language Proficiency Examination Services

State of Texas

County of Travis

The undersigned ÒAffiantÓ is a duly authorized representative of the Offeror for the purpose of making this Affidavit, and, after being first duly sworn, has deposed and stated and hereby deposes and states, to the best of his or her personal knowledge and belief as follows:

The term ÒOfferorÓas used herein, includes the individual or business entity submitting the Offer and for the purpose of this Affidavit includes the directors, officers, partners, managers, members, principals, owners, agents, representatives, employees, other parties in interest of the Offeror, and anyone or any entity acting for or on behalf of the Offeror, including a subcontractor in connection with this Offer.

1. **Anti-Collusion Statement.** The Offeror has not in any way directly or indirectly:

   a. colluded, conspired, or agreed with any other person, firm, corporation, Offeror or potential Offeror to the amount of this Offer or the terms or conditions of this Offer.

   b. paid or agreed to pay any other person, firm, corporation Offeror or potential Offeror any money or anything of value in return for assistance in procuring or attempting to procure a contract or in return for establishing the prices in the attached Offer or the Offer of any other Offeror.

2. **Preparation of Solicitation and Contract Documents.** The Offeror has not received any compensation or a promise of compensation for participating in the preparation or development of the underlying Solicitation or Contract documents. In addition, the Offeror has not otherwise participated in the preparation or development of the underlying Solicitation or Contract documents, except to the extent of any comments or questions and responses in the solicitation process, which are available to all Offerors, so as to have an unfair advantage over other Offerors, provided that the Offeror may have provided relevant product or process information to a consultant in the normal course of its business.

3. **Participation in Decision Making Process.** The Offeror has not participated in the evaluation of Offers or other decision making process for this Solicitation, and, if Offeror is awarded a Contract hereunder, no individual, agent, representative, consultant, subcontractor, or subconsultant associated with Offeror, who may have been involved in the evaluation or other decision making process for this Solicitation, will have any direct or indirect financial interest in the Contract, provided that the Offeror may have provided relevant product or process information to a consultant in the normal course of its business.

4. **Present Knowledge.** Offeror is not presently aware of any potential or actual conflicts of interest regarding this Solicitation, which either enabled Offeror to obtain an advantage over other Offerors or would prevent Offeror from advancing the best interests of the City in the course of the performance of the Contract.

5. **City Code.** As provided in Sections 2-7-61 through 2-7-65 of the City Code, no individual with a substantial interest in Offeror is a City official or employee or is related to any City official or employee within the first or second degree of consanguinity or affinity.

6. **Chapter 176 Conflict of Interest Disclosure.** In accordance with Chapter 176 of the Texas Local Government Code, the Offeror:

   a. does not have an employment or other business relationship with any local government officer of the City or a family member of that officer that results in the officer or family member receiving taxable income;

CITY OF AUSTIN
NON-COLLUSION,
NON-CONFLICT OF INTEREST, AND ANTI-LOBBYING AFFIDAVIT

b. has not given a local government officer of the City one or more gifts, other than gifts of food, lodging, transportation, or entertainment accepted as a guest, that have an aggregate value of more than $250 in the twelve month period preceding the date the officer becomes aware of the execution of the Contract or that OWNER is considering doing business with the Offeror.

c. as required by Chapter 176 of the Texas Local Government Code, Offeror must file a Conflict of Interest Questionnaire with the Office of the City Clerk no later than 5:00 P.M. on the seventh (7) business day after the commencement of contract discussions or negotiations with the City or the submission of an Offer, or other writing related to a potential Contract with the City. The questionnaire is available on line at the following website for the City Clerk:

http://www.austintexas.gov/department/conflict-interest-questionnaire

There are statutory penalties for failure to comply with Chapter 176.

If the Offeror cannot affirmatively swear and subscribe to the forgoing statements, the Offeror shall provide a detailed written explanation in the space provided below or, as necessary, on separate pages to be annexed hereto.

Offeror's Explanation:

7. **Anti-Lobbying Ordinance.** As set forth in the Solicitation Instructions, Section 0200, paragraph 7N, between the date that the Solicitation was issued and the date of full execution of the Contract, Offeror has not made and will not make a representation to a City official or to a City employee, other than the Authorized Contact Person for the Solicitation, except as permitted by the Ordinance.

Contractor's Name:	Berlitz Languages, Inc.
Printed Name:	Kristen Erdem
Title:	Business Development Manager for Texas and Oklahoma

Kristen Erdem

Signature of Officer or Authorized Representative:

Subscribed and sworn to before me this 12th day of October, 20 12.

Notary Public

My Commission Expires 7/31/16

EMILY GRIMES
MY COMMISSION EXPIRES
July 31, 2016

O.6 City of Austin, Texas Living Wages and Benefits Contractor Certifications

CITY OF AUSTIN, TEXAS
LIVING WAGES AND BENEFITS CONTRACTOR CERTIFICATION
(Please duplicate as needed)

SOLICITATION NO. JSD0129

Pursuant to the Living Wages and Benefits provision (reference Section 0400, Supplemental Purchase Provisions) the Contractor is required to pay to all employees directly assigned to this City contract a minimum Living Wage equal to or greater than $11.00 per hour.

I hereby certify under penalty of perjury that all of the below listed employees of the Contractor who are directly assigned to this contract:

(1) are compensated at wage rates equal to or greater than $11.00 per hour; and

(2) are offered a health care plan with optional family coverage.

(To add additional employees to this page, click the Add Button.)

	Employee Name	Employee Job Title
Add Delete	Lauren Kennedy	Director of Test Development

(3) all future employees assigned to this Contract will be paid a minimum Living Wage equal to or greater than $11.00 per hour and offered a health care plan with optional family coverage.

(4) Our firm will not retaliate against any employee claiming non-compliance with the Living Wage provision.

A Contractor who violates this Living Wage provision shall pay each employee affected the amount of the deficiency for each day the violation continues. Willful or repeated violations of the provision may result in termination of this Contract for Cause and subject the firm to possible suspension or debarment.

Contractor's Name:	Berlitz Languages, Inc.		
Signature of Officer or Authorized Representative:	Kristen Erdem	Date:	10/11/12
Printed Name:	Kristen Erdem		
Title:	Business Development Manager for Texas and Oklahoma		

Berlitz Testing

CITY OF AUSTIN, TEXAS
LIVING WAGES AND BENEFITS EMPLOYEE CERTIFICATION

Contract Number:	Solicitation No. JSD0129
Description of Services:	Language Proficiency Examination Services
Contractor Name:	Berlitz Languages, Inc.

Pursuant to the Living Wages and Benefits provision of the contract (reference Section 0400, Supplemental Purchase Provisions), the Contractor is required to pay to all employees directly assigned to this City contract a minimum Living Wage equal to or greater than $11.00 per hour. In addition, employees are required to certify that they are compensated in accordance with the Living Wage provision. Contractors are prohibited from retaliating against any employee claiming non-compliance with the Living Wage provision.

I hereby certify under penalty of perjury that I am directly assigned to this contract and that I am:

(1) compensated at wage rates equal to or greater than $11.00 per hour; and
(2) offered a health care plan with optional family coverage.

Employee's Title:	Director of Test Development		
Signature of Employee:	[signature]	Date:	10/11/12
Employee's Printed Name:	Lauren Kennedy		

Kristen Erdem
(Witness Signature)

KRISTEN ERDEM
(Printed Name)

CITY OF AUSTIN, TEXAS
LIVING WAGES AND BENEFITS CONTRACTOR CERTIFICATION
(Please duplicate as needed)

SOLICITATION NO. JSD0129

Pursuant to the Living Wages and Benefits provision (reference Section 0400, Supplemental Purchase Provisions) the Contractor is required to pay to all employees directly assigned to this City contract a minimum Living Wage equal to or greater than $11.00 per hour.

I hereby certify under penalty of perjury that all of the below listed employees of the Contractor who are directly assigned to this contract:

(1) are compensated at wage rates equal to or greater than $11.00 per hour; and

(2) are offered a health care plan with optional family coverage.

(To add additional employees to this page, click the Add Button.)

	Employee Name	Employee Job Title
Add Delete	Kristen Erdem	Business Development Manager for Texas and Oklahoma

(3) all future employees assigned to this Contract will be paid a minimum Living Wage equal to or greater than $11.00 per hour and offered a health care plan with optional family coverage.

(4) Our firm will not retaliate against any employee claiming non-compliance with the Living Wage provision.

A Contractor who violates this Living Wage provision shall pay each employee affected the amount of the deficiency for each day the violation continues. Willful or repeated violations of the provision may result in termination of this Contract for Cause and subject the firm to possible suspension or debarment.

Contractor's Name:	Berlitz Languages, Inc.		
Signature of Officer or Authorized Representative:	Kristen Erdem	Date:	10/11/12
Printed Name:	Kristen Erdem		
Title:	Business Development Manager for Texas and Oklahoma		

CITY OF AUSTIN, TEXAS
LIVING WAGES AND BENEFITS EMPLOYEE CERTIFICATION

Contract Number:	Solicitation No. JSD0129
Description of Services:	Language Proficiency Examination Services
Contractor Name:	Berlitz Languages, Inc.

Pursuant to the Living Wages and Benefits provision of the contract (reference Section 0400, Supplemental Purchase Provisions), the Contractor is required to pay to all employees directly assigned to this City contract a minimum Living Wage equal to or greater than $11.00 per hour. In addition, employees are required to certify that they are compensated in accordance with the Living Wage provision. Contractors are prohibited from retaliating against any employee claiming non-compliance with the Living Wage provision.

I hereby certify under penalty of perjury that I am directly assigned to this contract and that I am:

(1) compensated at wage rates equal to or greater than $11.00 per hour; and
(2) offered a health care plan with optional family coverage.

Employee's Title:	Business Development Manager for Texas and Oklahoma		
Signature of Employee:	Kristen Erdem	Date:	10/11/12
Employee's Printed Name:	Kristen Erdem		

(Witness Signature)

Jeffrey Price
(Printed Name)

Berlitz Testing

CITY OF AUSTIN, TEXAS
LIVING WAGES AND BENEFITS CONTRACTOR CERTIFICATION
(Please duplicate as needed)

SOLICITATION NO. JSD0129

Pursuant to the Living Wages and Benefits provision (reference Section 0400, Supplemental Purchase Provisions) the Contractor is required to pay to all employees directly assigned to this City contract a minimum Living Wage equal to or greater than $11.00 per hour.

I hereby certify under penalty of perjury that all of the below listed employees of the Contractor who are directly assigned to this contract:

(1) are compensated at wage rates equal to or greater than $11.00 per hour; and

(2) are offered a health care plan with optional family coverage.

(To add additional employees to this page, click the Add Button.)

	Employee Name	Employee Job Title
Add Delete	Alexis Nguyen	Testing Center Manager

(3) all future employees assigned to this Contract will be paid a minimum Living Wage equal to or greater than $11.00 per hour and offered a health care plan with optional family coverage.

(4) Our firm will not retaliate against any employee claiming non-compliance with the Living Wage provision.

A Contractor who violates this Living Wage provision shall pay each employee affected the amount of the deficiency for each day the violation continues. Willful or repeated violations of the provision may result in termination of this Contract for Cause and subject the firm to possible suspension or debarment.

Contractor's Name:	Berlitz Languages, Inc.		
Signature of Officer or Authorized Representative:	Kristen Erdem	Date:	10/11/12
Printed Name:	Kristen Erdem		
Title:	Business Development Manager for Texas and Oklahoma		

Berlitz Testing

CITY OF AUSTIN, TEXAS
LIVING WAGES AND BENEFITS EMPLOYEE CERTIFICATION

Contract Number:	Solicitation No. JSD0129
Description of Services:	Language Proficiency Examination Services
Contractor Name:	Berlitz Languages, Inc.

Pursuant to the Living Wages and Benefits provision of the contract (reference Section 0400, Supplemental Purchase Provisions), the Contractor is required to pay to all employees directly assigned to this City contract a minimum Living Wage equal to or greater than $11.00 per hour. In addition, employees are required to certify that they are compensated in accordance with the Living Wage provision. Contractors are prohibited from retaliating against any employee claiming non-compliance with the Living Wage provision.

I hereby certify under penalty of perjury that I am directly assigned to this contract and that I am:

(1) compensated at wage rates equal to or greater than $11.00 per hour, and
(2) offered a health care plan with optional family coverage.

Employee's Title:	Testing Center Manager		
Signature of Employee:		Date:	10/11/12
Employee's Printed Name:	Alexis Nguyen		

____
(Witness Signature)

Kristen Erden
____
(Printed Name)

O.7 City of Austin, Nonresident Bidder Provisions

City of Austin, Texas
NONRESIDENT BIDDER PROVISIONS
SOLICITATION NO. USD0129

A. Bidder must answer the following questions in accordance with Vernon's Texas Statutes and Codes Annotated Government Code 2252.002, as amended.

Is the Bidder that is making and submitting this Bid a "Resident Bidder" or a "Non-resident Bidder"?

- ○ Texas Resident Bidder - A Bidder whose principal place of business is in Texas and includes a Contractor whose ultimate parent company or majority owner has its principal place of business in Texas.
- ⦿ Non-resident Bidder

Bidder's Name:	Berlitz Languages, Inc.		
Signature of Officer or Authorized Representative:	Kristen Erdem	Date:	10/11/12
Printed Name:	Kristen Erdem		
Title:	Business Development Manager for Texas and Oklahoma		

O.8 City of Austin Local Business Presence Identification Form

City of Austin
Purchasing Office
Local Business Presence Identification Form
Request for Proposal JSD0129 – Language Proficiency Exam Services

A firm (Offeror or Subcontractor) is considered to have a Local Business Presence if the firm is headquartered in the Austin Corporate City Limits, or has a branch office located in the Austin Corporate City Limits in operation for the last five (5) years. The City defines headquarters as the administrative center where most of the important functions and full responsibility for managing and coordinating the business activities of the firm are located. The City defines branch office as a smaller, remotely located office that is separate from a firm's headquarters that offers the services requested and required under this solicitation.

OFFEROR MUST SUBMIT THE FOLLOWING INFORMATION FOR EACH LOCAL BUSINESS (INCLUDING THE OFFEROR, IF APPLICABLE).

NOTE: ALL FIRMS MUST BE IDENTIFIED ON THE MBE/WBE COMPLIANCE PLAN OR NO GOALS UTILIZATION PLAN. SECTION 0900 OF THE SOLICITATION.

USE ADDITIONAL PAGES AS NECESSARY

OFFEROR:

Name of Local Firm	Berlitz Languages, Inc.					
Physical Address	7000 North MoPac Expwy, Suite 200, Austin, Texas 78733					
Is Firm located in the Corporate City Limits? (circle one)	Yes			No		
In business at this location for past 5 yrs?	Yes			No		
Location Type:	Headquarters	Yes	No	Branch	Yes	No

SUBCONTRACTOR(S):

Name of Local Firm						
Physical Address						
Is Firm located in the Corporate City Limits? (circle one)	Yes			No		
In business at this location for past 5 yrs?	Yes			No		
Location Type:	Headquarters	Yes	No	Branch	Yes	No

SUBCONTRACTOR(S):

Name of Local Firm						
Physical Address						
Is Firm located in the Corporate City Limits? (circle one)	Yes			No		
In business at this location for past 5 yrs?	Yes			No		
Location Type:	Headquarters	Yes	No	Branch	Yes	No

City of Austin
Purchasing Office
Local Business Presence Identification Form
Request for Proposal JSD0129 – Language Proficiency Exam Services

---

ACKNOWLEDGEMENT

THE STATE OF TEXAS

COUNTY OF TRAVIS

I certify that my responses and the information provided on **Form 0605** are true and correct to the best of my personal knowledge and belief and that I have made no willful misrepresentations in this Section, nor have I withheld any relevant information in my statements and answers to questions. I am aware that any information given by me in this Section may be investigated and I hereby give my full permission for any such investigation and I fully acknowledge that any misrepresentations or omissions in my responses and information may cause my offer to be rejected.

OFFEROR'S FULL NAME AND ENTITY STATUS:

Kristen Erdem, Berlitz Languages, Inc.

Kristen Erdem

Signature, Authorized Representative of Offeror

Business Development Manager for Texas and Oklahoma

Title

10/11/12

Date

END

O.9 Minority- and Women-Owned Business Enterprise Procurement Program No Goals Form

**MINORITY- AND WOMEN-OWNED BUSINESS ENTERPRISE (MBE/WBE)
PROCUREMENT PROGRAM
NO GOALS FORM**

SOLICITATION NUMBER: JSD0129

PROJECT NAME: Language Proficiency Exam Services

The City of Austin has determined that no goals are appropriate for this project. Even though no goals have been established for this solicitation, the Bidder/Proposer is required to comply with the City's MBE/WBE Procurement Program, if areas of subcontracting are identified.

If any service is needed to perform the Contract and the Bidder/Proposer does not perform the service with its own workforce or if supplies or materials are required and the Bidder/Proposer does not have the supplies or materials in its inventory, the Bidder/Proposer shall contact the Small and Minority Business Resources Department (SMBR) at (512) 974-7600 to obtain a list of MBE and WBE firms available to perform the service or provide the supplies or materials. The Bidder/Proposer must also make a Good Faith Effort to use available MBE and WBE firms. Good Faith Efforts include but are not limited to contacting the listed MBE and WBE firms to solicit their interest in performing on the Contract; using MBE and WBE firms that have shown an interest, meet qualifications, and are competitive in the market; and documenting the results of the contacts.

Will subcontractors or sub-consultants or suppliers be used to perform portions of this Contract?

No X **If no, please sign the No Goals Form and submit it with your Bid/Proposal in a sealed envelope.**

Yes_____ **If yes, please contact SMBR to obtain further instructions and an availability list and perform Good Faith Efforts. Complete and submit the No Goals Form and the No Goals Utilization Plan with your Bid/Proposal in a sealed envelope.**

After Contract award, if your firm subcontracts any portion of the Contract, it is a requirement to complete Good Faith Efforts and the No Goals Utilization Plan, listing any subcontractor, subconsultant, or supplier. Return the completed Plan to the Project Manager or the Contract Manager.

I understand that even though no goals have been established, I must comply with the City's MBE/WBE Procurement Program if subcontracting areas are identified. I agree that this No Goals Form and No Goals Utilization Plan shall become a part of my Contract with the City of Austin.

Berlitz Languages, Inc.

Company Name

Kristen Erdem, Business Development Manager for Texas and Oklahoma

Name and Title of Authorized Representative (Print or Type)

Kristen Erdem 10/11/12

Signature **Date**

JSD0129 - Section 0900 No Goals

EXHIBIT A

Berlitz Proficiency Interview
Non-releasable Test Specification Version 1.0

For

Berlitz International

Prepared by

The University of Illinois at Urbana-Champaign (UIUC)
Foreign Language Assessment Group (FLAG)

Principal Investigator: Prof. Fred Davidson (DEIL)
Kadeessa Abdul-Kadir (Educational Psychology)
So-Young Jang (Educational Psychology)
Tammy Hsu (Curriculum & Instruction)
Youngshin Chi (Educational Psychology)

May 30, 2007

EXHIBIT C

Berlitz Proficiency Interview Validation Study Two

For

Berlitz International Inc.

Prepared by

The University of Illinois at Urbana-Champaign (UIUC)
Foreign Language Assessment Group (FLAG)

Principal Investigator: Prof. Fred Davidson (DEIL)
Kadeessa Abdul-Kadir (Educational Psychology)
So-Young Jang (Educational Psychology)
Tammy Hsu (Curriculum & Instruction)
Youngshin Chi (Educational Psychology)
Jiyoung, Kim (Educational Psychology)
Cary Lin (DEIL)

June 24, 2008

TABLE OF CONTENTS

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

EXECUTIVE SUMMARY

Background

In May of 2007, the UIUC FLAG prepared a validation report about the Berlitz Proficiency Interview (BPI). Following modern validity theory (Messick, 1989; Kane, 2006), that report (and this) provides arguments that score inferences made from the test are valid for the intended purpose. The elements of the validation argument in the 2007 report still hold, and they are as follows (quoting from the executive summary of that report):

1. that the nature of the BPI construct as in the test specifications is reflected adequately in the analysis of the speech samples
2. that the BPI is adaptive on two different levels as explained in the test specifications is reflected adequately in the speech samples. These features include (i) linking BPI levels of expectancy to the CEFR, (ii) linking raters' questions to examinees' context and responses.
3. that the four criteria used in the scoring procedures as in the test specifications provide adequate information about BPI's internal consistency.
4. that the training materials provide adequate instruction and information for the raters to conduct the BPI appropriately and effectively.
5. that the BPI is empirically derived and based on test specification development theories.
6. that the BPI is a reliable measure of speaking ability

These points of argument are set forth in greater detail in the 2007 report, and relevant discussion for each point is presented there. The above validity points have guided subsequent work to obtain evidence not available in 2007. This present report examines point (6), regarding reliability. In particular, the present report presents fully-crossed multiple rater analysis, in which each rating sample was judged by each rater. This allows a much fuller and richer analysis of rater agreement than in 2007, when we did pairwise double-rating analysis.

Findings

Reliability is consistency in test scoring – it is a matter of degree. How consistent is the testing across opportunities for inconsistency? In the BPI, the most logical source of inconsistency is disagreement among raters. If raters disagree, the next logical question is: why? On what parts or elements of the rating process?

The overall Cronbach's Alpha reliability in 2007 was .84 and is .86 in this present report, values which we take to be strong indicators of internal consistency of the rating process.[1] We present the relevant tables again, here in this summary; in all tables re-produced from this report or from the 2007 report here in the Exec Summary, relevant values are highlighted in yellow.

Table 10

*Reliability Analysis of Assessment Criteria (2007)**

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
Grammatical Accuracy	11.98	5.94	.61	.37	.82
Linguistic Range	11.51	4.84	.78	.62	.74
Phonological Control	12.08	5.29	.66	.47	.80
Fluency	11.76	5.31	.64	.48	.81
Overall					.84

* "Scale" refers to the grand or total mean if the specific criterion is omitted. For instance, the mean of 11.98 does not include Grammatical Accuracy.

[1] Ref. pg. 17 in the 2007 report and Table 15 in this present report, respectively.

Table 15

Reliability Analysis of Assessment Criteria (2008)

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
Grammatical Accuracy	11.50	3.39	.70	.49	.83
Linguistic Range	11.17	3.44	.76	.61	.80
Phonological Control	11.67	3.54	.62	.38	.86
Fluency	11.27	3.13	.76	.61	.80
Overall					.86

In both 2007 and 2008, we also compared raters directly as shown in the next tables. For 2007, the inter-rater correlations ranged from a low of .44 to a moderate value of .76 with an average (after transformation) of .59 as shown in Table 12, reproduced here.[2] In 2008, the same analysis revealed correlations that ranged from .48 to .81 showing again a moderate consistency between the paired raters with an average of 0.66 (after transformation) for all the possible 28 pairs of raters. Comparing this average with that of 2007 (Table 12 in this summary) which was .59, this showed a marked increase in the inter-rater reliability estimates when averaging across all pairwise comparisons.

Table 12

Pearson Correlations, Percent Agreement and Inter-rater Reliability Estimates (2007)

Rater	N	Percent agreement between paired raters (%)	Pearson Correlation	SEM	Reliability Estimates (after transformation)
R1 & R2	20	55.00	.39	.35	.54
R2 & R3	20	55.00	.64	.31	.70
R3 & R4	20	75.00	.57	.32	.66
R4 & R5	20	55.00	.75	.36	.76
R5 & R6	20	45.00	.67	.32	.72
R6 & R7	20	30.00	.30	.60	.44
R7 & R1	20	55.00	.37	.31	.51
Overall	140	53.00	.53	.37	.59

Table 11

Pairwise Correlations between Raters (After Z Transformation)(2008

	Mater R	R1	R2	R3	R4	R5	R6	R7
Mater R_Overall score	1							
R1_Overall score	.80	1						
R2_Overall score	.71	.71	1					
R3_Overall score	.60	.66	.71	1				
R4_Overall score	.71	.70	.61	.59	1			
R5_Overall score	.64	.64	.68	.76	.70	1		
R6_Overall score	.62	.61	.48	.53	.62	.60	1	
R7_Overall score	.80	.81	.73	.66	.70	.63	.53	1

Note. Average pairwise correlations for all 28 pairs is 0.66

Of greater import are the 2008 report's coefficients between the master rater and the rest of the raters, an analysis that was not feasible under the 2007 methodology. These comparisons with the master rater ranged from .62 to .80 and which (after transformation) present an average correlation of .72, in Table 9 in the present report, also reproduced here[3]. The analysis by comparison with the master rater is more important because it represents a typical method by which a testing system calibrates newly-hired rater/trainees: the veteran rater(s) serve as a touchstone during the training process.

Table 9.

*Paired correlation with the Master rater (2008)**

	Master Rater	
	Pearson' R(before transformation)	Pearson's R (after transformation)
R1	.832(**)	.80
R2	.659(**)	.71
R3	.483(**)	.60
R4	.665(**)	.72
R5	.541(**)	.64
R6	.523(**)	.62
R7	.841(**)	.80
Average		.72

[2] Ref. Table 12 in 2007 report.

[3] Ref. Table 9 in this present report.

All of these values compare quite favorably to industry-reported values; this is particularly true of our overall Cronbach's Alpha reliabilities of .84 (2007) and .86 (2008) cited above, given that the values obtained elsewhere in the industry often involve tests that contain some or solely objectively scored tasks.[4]

In summary, the next table (E-1) reports the overall comparison of the 2007 and 2008 analyses:

Table E-1

Comparison of the 2007 and 2008 analyses

	2007	2008
Number of Examinees	140	40
Number of Raters	7 (pairwise)	8 (fully crossed)
Average Reliability	.59 (average of all pairwise rater comparisons)	.66 (average of all pairwise rater comparisons) .72 (comparison with master rater)

In this present report, we calculated a simple percent agreement elsewhere in this study, and obtained an overall value of 58.3%.[5] There is no direct analogue to the percent agreement statistic in the 2007 report, because there were no multiple-rated data in 2007; however, the average 2007 percent agreement (across all rater pairs) was 53%, which is a rough comparison to the 2008 value.

---

[4] Ref. Appendix One in this present report. Objectively scored tasks (e.g. m/c items) usually provide higher reliabilities. The point being made here is that overall indexes in the .84-.86 range are very favorable for the BPI given that it is (a) a test under development, and (b) fully a rated exam with no objectively scored elements.

[5] Ref: Table 16 in this present report. Percent agreement is typically lower than reliability coefficients. It is possible to have perfect reliability coefficients but a non-perfect percent agreement, because the reliability values are comparisons of similar ranking, not absolute score values awarded. Note that industry standard (generally speaking) is to report reliability coefficients and not percent agreement, perhaps for this very reason.

Table 16

Rater Severity Estimates (2008)

Rater	Severity Estimate*	S.E	Infit**	Z-score
1	.56	.16	.90	-0.9
2	.53	.16	.76	-2.2
3	.08	.16	1.38	2.9
4	.71	.16	.78	-2.0
5	.13	.16	1.17	1.3
6	.43	.16	1.20	1.6
7	.06	.16	.74	-2.4
8	-.26	.16	1.06	0.5
Mean	.32	.16	1.00	
SD	.34	.00	.24	

Note: Fixed chi-square .38.1, df=7 at p=.00, Agreement = 58.3%. This chi-square indicates that raters differ in severity.

We next examined what aspects of the rating procedure contribute to the reliability results. These next analyses sought the answer to this question: what aspects of rater performance could be improved to enhance both the reliability values and the percent agreement? These analyses used two approaches, each named for the software employed: FACETS and GENOVA. The overall scale bracketed the data nicely.[6]

Table 18

Estimation of rating scales by FACETS

Proficiency level	Used (count)	%	Average Measure	Outfit Mean- square
2(A2)	19	1%	-4.06	1.1
3(B1)	410	32%	-1.90	0.9
4(B2)	693	54%	.05	1.1
5(C1)	123	10%	2.64	0.9
6(C2)	35	3%	3.85	1.4

[6] Ref. Figure 1, the 'scale' column, and Table 16 in this present report.

* Positive values indicate severity, whereas negative values indicate leniency. See the FACETS map, below, for more details.

** An infit value close to 1.0 is ideal for rater agreement.

Figure 1. FACETS MAP

```
-----------------------------------------
|Measr|+examinee|-rater |-criteria[7]|Scale|
-----------------------------------------
+   5 +         +       +         + (6) +
|     | *       |       |         |     |
|     |         |       |         |     |
|     | *       |       |         |     |
+   4 +         +       +         +     +
|     |  ABLE   |SEVERE |DIFFICULT|  5  |
|     |         |       |         |     |
|     | *       |       |         |     |
+   3 +         +       +         +     +
|     | *       |       |         | --- |
|     | *       |       |         |     |
|     | *       |       |         |     |
+   2 + **      +       +         +     +
|     |         |       |         |     |
|     |         |       |         |     |
|     |         |       |         |     |
+   1 +         +       + 3       +     +
|     | **      | 4     |         |  4  |
|     | **      | 1 2 6 | 1       |     |
|     | ***     | 5     |         |     |
*   0 * ****    * 3 7   *         *     *
|     | ******* | 8     |         |     |
|     | ****    |       | 4       |     |
|     |         |       |         |     |
+  -1 + *       +       + 2       +     +
|     | ***     |       |         |     |
|     |         |       |         | --- |
|     |LESS ABLE|LENIENT| EASY    |     |
+  -2 + *       +       +         +     +
|     |         |       |         |     |
|     | *       |       |         |     |
|     |         |       |         |     |
+  -3 + **      +       +         +     +
|     |         |       |         |     |
|     |         |       |         |     |
|     | *       |       |         |  3  |
+  -4 +         +       +         +     +
|     |         |       |         |     |
|     | *       |       |         |     |
|     |         |       |         |     |
+  -5 +         +       +         + (2) +
-----------------------------------------
|Measr| * = 1   |-rater |-criteria|Scale|
-----------------------------------------
```

[7] Criterion 1 refers to Grammatical Accuracy, criterion 2 refers to Linguistic Range, criterion 3 refers to Phonological Control, and criterion 4 refers to Fluency.

However, we detected some particular aspects of the rating in which raters disagreed with each other and/or appeared to need further clarification of the rating scale. These matters can be resolved by further training and are dealt with in greater detail in the report, below.

Hence, our recommendations here are as follows, copied from the conclusion of this present report:

1) Train[8] raters to understand and use the assessment criteria effectively by providing more standardization sessions. The standardization sessions should provide avenues for raters to give and receive feedback as well as to allow for introspection of their own rating experiences. These sessions should be conducted fairly regularly.

2) Multiple-score occasionally BPI test administration to monitor rater consistency and inter-reliability.[9]

3) Increase rater reliability by introducing a new rating model which includes an extra rater. The role of the first rater is to conduct the interview and rate the examinee. The role of the second rater is to rate the examinee. This new model is to be used for each or selected live administration of the BPI.

4) Fully-crossed comparisons (as in this 2008 report) remain the best way to analyze rater reliability.

---

[8] This should really say 'Continue to train…'; the point being made is that the training is clearly having an effect: overall reliabilities in the range of .84 to .86 indicate that to be the case.

[9] This is probably the single most important recommendation – ongoing quality control monitoring will ensure that the solid measurement qualities reported here do continue, and it will also allow us to check if changes to the training procedure (ref. recommendation (1) above) have an impact.

1. INTRODUCTION

The BPI validation study is an ongoing process which began in the test's pre-operational to operational stages in an attempt to constantly evaluate the quality of the test so as to improve its validity and reliability. Based on the findings from the first validation report, which indicated a gap between the groups of raters in their inter-rater reliability, the second validation study was initiated in Fall 2007 to gather evidence and further investigate raters' performance and the extent to which their ratings are consistent.

The current study employs different statistical approaches, described more fully below, to examine rater performance. Multiple instruments will first be used to demonstrate the different aspects of the findings and then synthesized to conform or refute these findings, which will thus enhance the credibility and comprehensiveness of the validity claims.

2. SOURCE OF DATA AND ANALYSIS

Data for this validation study comprised observed scores of 40 examinees, 8 raters including a master rater scored speech samples of all 40 examinees. In Fall 2007, a master rater was selected by FLAG based on analysis of raters' interview skills and their grading patterns. The master rater selected had interviewed and scored 40 examinees' performance. After finishing this first round, seven raters individually re-scored the same speech samples.

For this study, only quantitative methods were employed to analyze rating data, and evidence were collected and displayed using five different statistics with multiple purposes which included descriptives, correlation analysis, estimates of internal consistency, GENOVA, and multi-facets Rasch analysis(FACETS). Descriptives provide mean and standard deviations of ratings. Correlation analysis describes the degree of relationship between a rater and assessment criteria. GENOVA based on G-study estimates random variance components for multi-facets regarding examinees, raters and assessment criteria. FACETS analyses represent estimates of rater severity and intra-rater reliability. In addition, rater severity of the different assessment criteria is also provided by the RASCH model. Aside from the above, bias analysis using FACETS was also conducted to look into any interaction effects between rater and assessment criteria.

3. FINDINGS

3.1 Descriptives

3.1.1 Profile analysis of examinees

This section presents the results of the demographic and background information of the examinees. 62.5% of those who participated in the current study were male, while only 37.5% were females. The countries of origin for examinees include Philippines (52.5%), Argentina (22.5%), Mexico (17.5%) and India (7.5%).

3.1.2 Mean and SD for rating criteria

Table 1

Means and Standard Deviations for Each Assessment Criterion

Assessment Criteria	Minimum	Maximum	Mean	Std. Deviation
Average Score	2	6	3.75	.694
Grammatical Accuracy	2	6	3.71	.713
Linguistic Range	2	6	4.03	.659
Phonological Control	2	6	3.53	.721
Fluency	2	6	3.93	.756

Table 1 presents the results of the overall descriptive statistics for each rating criterion as well as the average score. The mean of the average score of the eight raters is 3.75 with Standard Deviation .70 which indicates the average proficiency level of examinees taking BPI would be between B1 and B2, which coincides with the independent user proficiency level according to the CEFR. In terms of the minimum and maximum scores, no examinee was in the A1 level, the basic user. As far as the assessment criteria are concerned, the mean for Linguistic Range is 4.03 with SD .66, indicating that examinees' scored highest on this criterion, whereas the mean for Phonological Control was the lowest among four criteria.

Table 2

Comparisons of Means and Standard Deviations of 8 Raters for Average Scores

Rater/ Average score	Range	Minimum	Maximum	Mean	Std. Deviation
Mater_ Average score	3	3	6	3.93	.656
R1_Average score	4	2	6	3.80	.723
R2_ Average score	2	3	5	3.63	.586
R3_ Average score	3	3	6	3.73	.905
R4_ Average score	3	2	5	3.73	.599
R5_ Average score	4	2	6	3.65	.864
R6_ Average score	2	3	5	3.68	.526
R7_ Average score	3	3	6	3.90	.591

Table 2 to Table 6 present the comparison of the mean and standard deviations of the master rater and the other seven raters for the average scores and each assessment criterion. Findings from Table 2 displays that of eight raters, the highest ratings for the overall scores were given by the Master Rater (Mean=3.93) followed by Rater 7(Mean=3.90), which seems to reflect they are relatively more lenient than the rest of raters. In terms of severity, Rater 3 and Rater 4 had similar severity level in rating for which the means for the overall BPI score are 3.73 respectively, and Rater 2 has lowest overall mean score among eight raters, indicating that person is the most stringent raters in this study.

Table 3

Comparisons of Means and Standard Deviations of 8 Raters for Grammatical Accuracy

Rater/ Criteria	Range	Minimum	Maximum	Mean	Std. Deviation
Master_GA	3	3	6	3.95	.597
R1_GA	4	2	6	3.80	.723
R2_GA	3	3	6	3.80	.648
R3_GA	3	3	6	3.65	.893
R4_GA	3	2	5	3.65	.622
R5_GA	3	3	6	3.63	.868
R6_GA	1	3	4	3.40	.496
R7_GA	3	3	6	3.78	.698

Table 3 displays the comparison of the mean scores and standard deviations of the master rater and the other seven raters for the Grammatical Accuracy assessment criterion.

Overall, the master rater gave the highest ratings to Grammatical Accuracy (Mean=3.95), whereas four other raters gave a score below the mean (See Table 1, Mean =3.71) of this assessment criterion. They were given by Rater 3(Mean=3.65), Rater 4(Mean=3.65), Rater 5(Mean=3.63), and Rater 6(Mean=3.40). Note also that the range used by Rater 6 for Grammatical Accuracy was between B1 and B2, suggesting that this rater used a limited range of the scoring option.

Table 4

Comparisons of Means and Standard Deviations of 8 Raters for Linguistic Range

Rater/ Criteria	Range	Minimum	Maximum	Mean	Std. Deviation
Master_LR	3	3	6	4.25	.776
R1_LR	4	2	6	3.88	.757
R2_LR	3	3	6	3.98	.577
R3_LR	3	3	6	4.20	.723
R4_LR	3	2	5	3.80	.564
R5_LR	4	2	6	4.05	.783
R6_LR	1	4	5	4.08	.267
R7_LR	3	3	6	4.05	.597

As for the Linguistic Range, ratings given by each rater were generally high. The mean rating by the master rater was the highest (Mean=4.25), compared to Rater 1(Mean=3.88) and Rater 4 (Mean=3.80). Again, Rater 6 used a very limited range of scoring option, whereas the proficiency levels were chosen between C2 and A2 for other raters. It is suggested that Rater 6's handling of Linguistic Range can be further improved.

Table 5

Comparisons of Means and Standard Deviations of 8 Raters for Phonological Control

Rater/ Criteria	Range	Minimum	Maximum	Mean	Std. Deviation
Master_PC	3	3	6	3.33	.764
R1_PC	3	2	5	3.35	.770
R2_PC	2	3	5	3.28	.506
R3_PC	3	3	6	3.63	.979
R4_PC	3	2	5	3.48	.599
R5_PC	3	3	6	3.53	.816
R6_PC	2	3	5	3.95	.316
R7_PC	2	3	5	3.73	.599

Table 5 shows the comparison of the mean scores and standard deviations for Phonological Control. Overall, the highest mean score was given by Rater 6 (Mean=3.95), while the lowest was from Rater 2 (Mean=3.28). On the whole, the eight raters used options between B1 and C2 for scoring this criterion.

Table 6

Comparisons of Means and Standard Deviations of 8 Raters for Fluency

Rater/Criteria	Range	Minimum	Maximum	Mean	Std. Deviation
Master_F	3	3	6	4.23	.800
R1_F	4	2	6	3.90	.841
R2_F	2	3	5	3.90	.545
R3_F	3	3	6	3.93	.944
R4_F	3	2	5	3.85	.622
R5_F	4	2	6	4.15	.802
R6_F	2	3	5	3.63	.586
R7_F	4	2	6	3.88	.723

As for Fluency, raters who gave the three highest mean scores were the master rater (Mean=4.23), Rater 5(Mean=4.15), and Rater 3(Mean=3.93), compared to Rater 6 who gave the lowest mean score (Mean=3.63). Rating pattern of Rater 2 and Rater 6 might show central tendency by using the limited rating scale.

3.2 Correlation Analysis

Correlation analyses were employed for analyzing the data using Pearson Product Moment Correlation. Reliability estimates between a master rater and individual raters using Fisher Z transformation (Lynch, 2003) were computed and the results[10] between the average score and each criterion are presented in Table 7.

Table 7.

Correlations between Average Score and Four Criteria

	Average Score	Grammatical Accuracy	Linguistic Lange	Phonological Control	Fluency
Grammatical Accuracy	.790(**)	1			
Linguistic Range	.759(**)	.635(**)	1		
Phonological Control	.690(**)	.530(**)	.555(**)	1	
Fluency	.786(**)	.642(**)	.740(**)	.550(**)	1

** Correlation is significant at the 0.01 level (2-tailed).

Looking at the correlation matrix in the off-diagonal part in the Table 7, all criteria were highly correlated with each other. The correlation between Grammatical Accuracy and average score is .79, indicating significantly high association. This indicates that the importance of Grammatical accuracy as an important criterion. Fluency and Linguistic range were highly correlated with the overall score, with coefficient values of .78, and .75 respectively. Comparing to the other correlation coefficients, Phonological control (.69) has the lowest association with the average score.

[10] Careful interpretation of the correlation matrix should be made since the correlation only refers to the degree of association between a dependent variable and independence variables, rather than cause and effect.

Table 8.

Correlations between Individual Average Score and Assessment Criteria across All Raters

	Master	R1	R2	R3	R4	R5	R6	R7
GA	.842(**)	.804(**)	.743(**)	.956(**)	.699(**)	.847(**)	.511(**)	.691(**)
LR	.743(**)	.889(**)	.655(**)	.830(**)	.896(**)	.785(**)	.361(*)	.815(**)
PC	.664(**)	.774(**)	.530(**)	.864(**)	.589(**)	.849(**)	.571(**)	.718(**)
F	.815(**)	.893(**)	.602(**)	.875(**)	.712(**)	.781(**)	.760(**)	.871(**)

The correlations between the master rater score on each criterion and the other raters' scores are presented in Table 8 in order to explore how raters perceive the four criteria. Each column reports the correlations between each rater's average score and each assessment criterion. It can be shown that Grammatical accuracy influences more on the average scores by the master rater (.842), compared to the other raters. As for the other raters, different criterion seems to have higher association with their own average scores.

Table 9.

Pairwise Correlations with the Master Rater

	Master Rater	
	Pearson' R(before transformation)	Pearson's R(after transformation)
R1	.832(**)	.80
R2	.659(**)	.71
R3	.483(**)	.60
R4	.665(**)	.72
R5	.541(**)	.64
R6	.523(**)	.62
R7	.841(**)	.80
Average		.72

Table 9 presents correlations between the Master rater and each individual rater.

Pearson correlation coefficients and Fisher Z transformation were used to get estimates of inter-rater reliability. The estimates of inter-rater reliability ranged between .60 and .80 The highest estimates of inter-rater reliability were between the Master rater and Rater 1 as well as the Master rater and Rater 7. The lowest estimate was between the Master rater and Rater 3. The average of the estimate was .72, which is moderately high for this group of raters. The agreement between the Master rater and the rest of the rater is approximately 52% (Rsquared of .72).

In addition to examining the correlations between the master rater and each individual rater, correlations between paired raters were also examined. Table 10 and 11 report the inter-rater correlations before and after Fisher Z transformation. As can be noted, the pairwise correlations between raters as presented in Table 11 suggest a moderate to high inter-rater correlations. These correlations ranged from .48 to .81 which showed a moderate consistency between the paired raters. The highest pairwise correlation was between R1 and R7 with a correlation of .81 whilst the lowest was between R2 and R 6 with a correlation of 0.48. The average correlation for the 28 pairs is .66.

Table 10

Pairwise Correlations between Raters (Before Z transformation)

	Mater R	R1	R2	R3	R4	R5	R6	R7
Mater R_Overall score	1							
R1_Overall score	.83	1						
R2_Overall score	.66	.67	1					
R3_Overall score	.48	.57	.67	1				
R4_Overall score	.67	.64	.50	.47	1			
R5_Overall score	.54	.54	.60	.76	.65	1		
R6_Overall score	.52	.50	.34	.40	.52	.48	1	
R7_Overall score	.84	.85	.70	.57	.64	.53	.39	1

Table 11

Pairwise Correlations between Raters (After Z Transformation)

	Mater R	R1	R2	R3	R4	R5	R6	R7
Mater R_Overall score	1							
R1_Overall score	.80	1						
R2_Overall score	.71	.71	1					
R3_Overall score	.60	.66	.71	1				
R4_Overall score	.71	.70	.61	.59	1			
R5_Overall score	.64	.64	.68	.76	.70	1		
R6_Overall score	.62	.61	.48	.53	.62	.60	1	
R7_Overall score	.80	.81	.73	.66	.70	.63	.53	1

Note. Average pairwise correlation for all 28 pairs is 0.66

3.3 The GENOVA Analysis

Generalizability study was conducted to investigate multiple sources of variability in the examinee's test scores. For the BPI, different sources of errors, such as raters and assessment criteria can have effect on the test scores. The theory is rooted in classical test theory (CTT) and analysis of variance (ANOVA). Generalizability theory considers both systematic and unsystematic sources of error variation and disentangles both error variations at once. Each characteristic of the measurement situation (e.g., item, test form) is called a facet and observations are defined by all possible combinations of the levels of the facets.

For the G-study, two facets considered for the analysis included the rater effect and the assessment criteria. The design for the G-study was two- facet P x R x I (person by rater by item).

Table 12

Variance Analysis of Two- Facet Model (P X R X I Design)

Source of Variance	Variance	Percent
p (persons)	0.2369	24.27
r (raters)	0.0130	1.33
i (items= assessment criteria)	0.0015	.15
pr (interaction between pr)	0.0229	2.35
pi (interaction between pi)	0.0115	1.18
ri (interaction between ri)	0.4817	49.34
pri, e	0.2087	21.38
Total	.9762	100

Table 12 reports the findings of the G-study. It can be noted that variance in performance can be explained largely by interaction of raters with items (criteria) as well as the different abilities of the examinees. It is found that examinee contributed 24% of the total variance whilst raters contributed only 1.33 %. 15 % of the magnitude of variance was contributed by the assessment criteria. These individual aspects would positively affect the scoring process. Interesting to note is the large amount of interaction between rater and assessment criteria. This constituted almost 50% of the total variance. This may suggest that the way the rater interacts with the assessment criteria can affect the scoring of BPI and hence the score that an examinee receives from a particular rater. High rater and item (criteria) interaction suggests that depending on which rater an examinee interviews, the rater will interact with the scoring criteria differently.

Table 13

Variance analysis of one-facet model (P X R Design)

Source of Variance	Variance	Percent
P (persons)	0.2087	9.57
R (raters)	1.7444	79.96
PR, e	0.2283	10.47
Total	2.1814	

This analysis (P X R design) was conducted based on the average ratings given by each rater (average scores). Table 13 shows that the largest magnitude of variance was due to raters (79.96%) and this indicates that the assignment of rater for each examinee affects the ranking of the examinee. Low magnitude of persons suggests that on the whole, there was not much difference in examinee ability. 10.47% of the variance can be accounted by the interaction between rater, examinees and residual.

Table 14

G- Coefficient based on D-study

Number of raters	Generalizability Coefficient[11]	Phi
8	.88	.46
6	.85	.39
4	.79	.42
2	.65	.17
1	.48	.09

[11] Perfect value for G-coefficient is closer to 1 as the same as the other reliability index.

Table 14 shows the Generalizability coefficients which indicate a reliability index produced by D-study. These findings provide information on the reliability of the BPI when different numbers of raters are used in the scoring of the assessment. As can be noted, a moderately high G-coefficient of .88 for the double-scoring data with 8 raters is reported for this analysis. This moderately high G-coefficient decreased substantially when the number of raters is reduced from 8 to the operational single rater situation. If we look at Table 14, the G-coefficient for the operational single rater situation was only .48. This is considered a low reliability estimate for the BPI.

3.4 Measures of Internal Consistency

Table 15

Reliability Analysis of Assessment Criteria

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
Grammatical Accuracy	11.50	3.39	.70	.49	.83
Linguistic Range	11.17	3.44	.76	.61	.80
Phonological Control	11.67	3.54	.62	.38	.86
Fluency	11.27	3.13	.76	.61	.80
Overall					.86

The Cronbach's Alpha as a measure of internal consistency of the BPI reports a value of .86 in Table 15 for the four criteria used in the scoring of the BPI. This relatively high value suggests that the criteria used for scoring can be considered reliable for measuring each of four areas namely grammatical accuracy, linguistic range, fluency and phonological control. However, a word of caution when using Cronbach's Alpha for reporting internal consistency of performance-based testing such as the BPI as the scores were obtained from rater judgments which involved subjective scoring rather than objective scoring.

3.5 FACETS Analysis

```
-----------------------------------------
|Measr|+examinee|-rater |-criteria|Scale|
-----------------------------------------
+   5 +         +       +         + (6) +
|     | *       |       |         |     |
|     |         |       |         |     |
|     | *       |       |         |     |
+   4 +         +       +         +     +
|     |         |       |         |  5  |
|     |         |       |         |     |
|     | *       |       |         |     |
+   3 +         +       +         +     +
|     | *       |       |         | --- |
|     | *       |       |         |     |
|     | *       |       |         |     |
+   2 + **      +       +         +     +
|     |         |       |         |     |
|     |         |       |         |     |
|     |         |       |         |     |
+   1 +         +       + 3       +     +
|     | **      | 4     |         |  4  |
|     | **      | 1 2 6 | 1       |     |
|     | ***     | 5     |         |     |
*   0 * ****    * 3 7   *         *     *
|     | ******* | 8     |         |     |
|     | ****    |       | 4       |     |
|     |         |       |         |     |
+  -1 + *       +       + 2       +     +
|     | ***     |       |         |     |
|     |         |       |         | --- |
|     |         |       |         |     |
+  -2 + *       +       +         +     +
|     |         |       |         |     |
|     | *       |       |         |     |
|     |         |       |         |     |
+  -3 + **      +       +         +     +
|     |         |       |         |     |
|     |         |       |         |     |
|     | *       |       |         |  3  |
+  -4 +         +       +         +     +
|     |         |       |         |     |
|     | *       |       |         |     |
|     |         |       |         |     |
+  -5 +         +       +         + (2) +
-----------------------------------------
|Measr| * = 1   |-rater |-criteria|Scale|
-----------------------------------------
```

Figure 1. FACETS MAP

Figure 1 shows the facet map for the BPI examinees, raters and the assessment criteria. Under the examinee column, one can note that the examinees on the BPI demonstrated a wide range of abilities. The spread of the examinees ranged from -5 to +5 on the logit scale although the majority of examinees was clustered around the 0 logit scale. On the other hand as reported for G-studies, the magnitude of variance contributed by examinees was small (10%) suggesting less variability in the examinees' performance. This seems to suggest that although FACETS shows wider ability on the logit scale, most of the ability levels were clustered around 0 on the logit scale.

The most able examinees are at the top of the scale whereas the least able at the bottom. The third column shows the severity variation among raters. The most severe rater was at the top and the least severe rater was at the bottom. A detail description of rater severity and consistency is reported in Table 16 As for the criteria in the rating scale, these were quite separated in their place on the logit scale. The criteria above 0 on the logit scale were considered to be more difficult and those below.

Table 16

Rater Severity Estimates

Rater	Severity Estimate	S.E	Infit	Z-score
1	.56	.16	.90	-0.9
2	.53	.16	.76	-2.2
3	.08	.16	1.38	2.9
4	.71	.16	.78	-2.0
5	.13	.16	1.17	1.3
6	.43	.16	1.20	1.6
7	.06	.16	.74	-2.4
8	-.26	.16	1.06	0.5
Mean	.32	.16	1.00	
SD	.34	.00	.24	

Note: Fixed chi-square .38.1, df= 7 at p=.00, Agreement = 58.3%

Table 16 presents the rater measurement report. As can be noted, the Chi-square value was significant which implied that the raters used for this double-rating study consistently differed from one another in their ratings. However, there was a moderate 58.3 percent agreement among the raters. The 8 raters who participated in this study ranged from -.26 to .71 in severity. As can be noted, on the average, the raters were tougher than easy on the examinees although these values were not that extreme. They were more close to 0 on the logit scale. The toughest was Rater 4 (.71) and the most lenient was Rater 8, the master rater. But if we note the fit values for the different raters, Rater 8 which was the master rater had almost a perfect value of 1.06 which suggests that she was very consistent in her rating (reliable) although the most lenient. If we use recommended standard, which is [1.00 ± (.24 x 2 = .48)], then no raters were found to be misfitting as they were within the acceptable range that is between 1.48 and .52 which means that all raters were self-consistent in their own ratings. However, if we adopt a more conservative or strict measure recommended by McNamara (1996) for misfitting, [0.8-1.2], Raters 2, 3, 4, 6, and 7 would be considered inconsistent in their own ratings.

Table 17

Criteria Measurement Report (arranged by 3N)

Criteria	Estimate	S.E	Infit	Z-score
GA (1)	.38	.11	1.07	-0.9
LR (2)	-.93	.11	.80	-2.3
PC (3)	1.07	.11	1.26	3.3
Fluency (4)	-.52	.11	.87	-1.5
Mean	.00	.11	1.00	
SD	.90	.11	.21	

Table 17 reports the absolute difficulty of the criteria in relation to the other facets. One can note that the raters seem to be more lenient with linguistic range (-.93) and fluency but stricter on phonological control (1.07) and grammatical accuracy (.38). Phonological control was the most severely rated or in other words examinees scored lowest on this criterion as compared to the other criteria. This has some implications as to which criteria the raters consider to be most important when assessing examinees on the BPI.

None of the criteria was found to be misfitting as they were within the acceptable range of .58 and 1.42, [1.00 ± (.21 x 2 = .42)]. However, if we consider the z-score, the Linguistic range and Phonological control have tendency towards influencing raters in giving inconsistent scores.

Table 18

Estimation of Rating Scales by FACETS

Proficiency level	Used (count)	%	Average Measure	Outfit Mean- square
2(A2)	19	1%	-4.06	1.1
3(B1)	410	32%	-1.90	0.9
4(B2)	693	54%	.05	1.1
5(C1)	123	10%	2.64	0.9
6(C2)	35	3%	3.85	1.4

Table 18 presents the estimates of BPI rating scale. In the first three columns, five proficiency levels were reflected in the analysis. The most frequently used proficiency level was B2 (54%) and the least was A2 (1%). The outfit mean squares provide an indicator of the quality of each proficiency level. As reported in Table 18, the outfit mean squares suggest that the CEF descriptors used for describing proficiency levels of examinee works well for the BPI assessment. Table 18 displays no noticeable misfit values for all the proficiency levels.

Table 19

Bias Analysis Between Raters and Assessment Criteria

Bias size	Model S.E	Rater	Severity(Rater)	Criteria	Difficulty
1.81*	.32	6	.43	Phonological control	1.07
.73 *	.31	5	.13	Fluency	-.52
.64 *	.31	1	.56	Grammatical Accuracy	.38
.62 *	.31	8	-.26	Fluency	-.52
-.85 *	.34	2	.53	Phonological Control	1.07
-1.06 *	.31	6	.43	Fluency	-.52
-1.09 *	.33	6	.43	Grammatical Accuracy	.38
-1.42 *	.33	8	-.26	Phonological Control	1.07

*Note.** Bias size significant at $p < .05$.

Table 19 reports the bias analysis between the raters and the assessment criteria. It can be noted that several of the raters involved in this study have tendency towards biasness in certain assessment criteria. Master rater 8 seems to exhibit biasness on Fluency and Phonological control. This suggests that some form of systematic anomaly occurs between the rater and the assessment criteria. Of all the eight raters, Rater 6 displayed the most frequent occurrence of biasness for three of the four assessment criteria. These three criteria where raters have the most difficulty include Phonological control, Fluency and Grammatical Accuracy.

Table 20

Bias Analysis Between Raters and Examinees' Gender and Language

Examinee				
Number	Bias Size	Gender	Language	Rater
1	-2.60,	Male	Spanish	6
4	1.71	Male	Tagalog	5
5	1.90, -2.51	Male	Spanish	3,7
12	1.80	Male	Tagalog	5
17	-2.51	Female	Tagalog	5
22	2.30	Female	Spanish	3
23	1.85	Female	Spanish	7
24	4.59,-2.47 ,3.42	Male	Spanish	3,4,5
25	-2.37, 4.75,-2.20,-2.47	Male	Spanish	2,3,4,6
28	2.10	Male	Hindi	8
30	1.80	Female	Tagalog	5
32	-2.61	Male	Tagalog	1
35	-3.89	Male	Spanish	1
37	3.50	Female	Spanish	6
38	-3.21, 2.67	Male	Spanish	1,6
39	2.09	Male	Spanish	2
40	-3.89, 2.30	Male	Spanish	1,6

Note. * Bias size significant at $p < .05$.

Table 20 shows the findings of the bias analysis between raters and examinees' gender and language. All bias size reported were significant at the 0.05 level. It can be noted rating is not affected by the gender of the examinees. However, the language spoken by the examinees seemed to affect raters' judgment. Occurrence of biasness was largely exhibited towards examines who spoke Spanish followed by Tagalog. It can also be noted that Rater 5 exhibited tendency of biasness to examinees who spoke Tagalog followed by Rater 6 who seemed to be biased towards Spanish speaking examinees.

4. DISCUSSION

The evidence presented in the findings lends some support to the reliability of the double-scoring data. The study findings suggest that BPI raters' performance have further improved compared to the findings of the first validation study. The multiple instruments were compared to look for evidences for convergence or divergence of the interpretations.

On the whole, the group of raters exhibited moderate estimates of inter-rater reliability between the Master rater and each rater (.60-.80); and pairwise comparisons between raters (.41-.81) which was supported by the results of FACETS analysis that indicated a moderate agreement between the Master rater and other raters (58.3%). In addition, the moderately high estimate of internal consistency of assessment criteria also suggests that the assessment criteria function well for the BPI. Beside estimates of inter-rater reliability, Generalizability analysis revealed that for this group of 8 raters, the G-coefficient was .87 which is comparable to estimate of rater internal consistency. Nevertheless, G-coefficient operational single rater was only. 47, which indicated low self-consistency of their own ratings. The G analysis also revealed that there was interaction between rater and assessment criteria when each of the assessment criteria was used for scoring decisions, and this seems to affect the performance of rating (variance-49.34%). However when the average score of each examinee was considered, the high magnitude of rater variance (79.96%) suggests that raters varied in their rating according to the examinees they interviewed. Answers to exactly how these raters varied in their own ratings were provided by the FACETS analysis. Firstly, raters showed less consistency within their own ratings. Several of the raters (Rater 2, 3, 4, and 7) would benefit from rater trainings in order to improve their own intra-rater consistency. In addition, from the bias analysis, it was also revealed that certain raters exhibited tendencies towards biasness in certain assessment criteria and examinee's language.

5. RECOMMENDATION

1) Train raters to understand and use the assessment criteria effectively by providing more standardization sessions. The standardization sessions should provide avenues for raters to give and receive feedback as well as to allow for introspection of their own rating experiences. These sessions should be conducted fairly regularly.

2) Multiple-score occasionally BPI test administration to monitor rater consistency and inter-reliability.

3) Increase rater reliability by introducing a new rating model which includes an extra rater. The role of the first rater is to conduct the interview and rate the examinee. The role of the second rater is to rate the examinee. This new model is to be used for each or selected live administration of the BPI.

6. CONCLUSION

The study was aimed at examining the quality of rating performed by the BPI raters. The findings from multiple analyses suggest that more efforts need to be undertaken in order to enhance the reliability of the BPI. These efforts include advocating a strong rater training program which focuses on improving raters understanding and use of the assessment criteria. Rater training should also reflect rater consistency in terms of rating severity addressing possible tendency towards biasness. In addition, Berlitz needs to consider introducing a new rating model which would allow monitoring of the operational single rater situation in order to further increase rater reliability. This will go a long way towards enhancing and improving the quality and credibility of the test.

REFERENCES

Lynch, B. K. (2003). Language assessment and program evaluation. Edinburgh: Edinburgh University Press.

McNamara, T. (1996). Measuring second language performance. New York: Longman.

Kane, M.T. (2006) Validation. In R. Brennan (Ed.), *Educational Measurement, Fourth Edition*. (Washington, DC: American Council on Education and National Council on Measurement in Education) pp. 17-64.

Messick, S. (1989) "Validity." In Linn, R. L. (Ed.) *Educational Measurement*. New York: Macmillan/American Council on Education, 13 - 103.

EXHIBIT D

BERLITZ TEST OF WRITING SKILLS (BTWS)
BERLITZ INTERNATIONAL, INCORPORATED

TEST SPECIFICATION
VERSION 1.00

FOREIGN LANGUAGE ASSESSMENT GROUP (FLAG)
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

May 21, 2009

TABLE OF CONTENTS

INTRODUCTION

The Berlitz Test of Writing Skill (BTWS) which assesses second language writing has been developed through the collaborative efforts of Berlitz International, Inc. and the University of Illinois at Urbana-Champaign. The test is designed to assess business and professional use of English writing in a workplace setting. Since its formulation in September 2008, three forms have been developed. This test is intended to be used by organizations, international corporations and government agencies throughout the world in their recruitment, staff training and development, and benchmarking processes.

The BTWS includes three tasks that were based on the outcomes of three activities: research on writing tests available on the market (see Appendix B), a needs analysis study (see Appendix C) and a pilot test (see Appendix D). Effort is made to select task types that best reflect the 'situational authenticity' (Bachman and Palmer, 1996), in the sense that the representative writing tasks in the real business setting are properly selected. The three chosen tasks are: write an e-mail, a memo and a report respectively. In 'Write an e-mail', test-takers are asked to write an e-mail reply according to a colleague's e-mail message. In 'Write a memo', test-takers need to write a short memo announcement to a wider audience in the company with regard to a newly launched policy. In 'Write a report', test-takers are expected to demonstrate their ability to write a professional evaluation report based on given graphs to potential or existing clients. Given a variety of factors that influence the level of difficulty in each task, the expected level of difficulty for 'Write an e-mail' is the easiest while 'Write a report' is the most difficult.

Findings from the needs analysis and pilot test also contributed to the development of the analytic rating scale (see Appendix E). Adapted from the Common European Framework of Reference scale, the BTWS rating scale has been continuously modified according to the results of pilot rating studies to better reflect the context of the BTWS.

This test specifications (specs) document serves as a blueprint to describe in detail the contents of the BTWS. The specifications provide information in relation to the different tasks of the test and detailed descriptions of the constructs; in addition, these specs provide all the rating materials necessary for operational testing. The three test forms have been transmitted to Berlitz as separate deliverables.

As an evolving instead of a fixed product, the BTWS spec has progressed and been refined continuously in the different stages of the project, in an effort to produce a defensible, theoretical solid and business-oriented writing test.

PART 1 : TASK TYPES

1. Item type 1: Write an e-mail

General Description (GD): Writing e-mails is found to be one of the most common written communications in the workplace settings. This portion of the test hence requires test-takers to demonstrate their ability to write a short e-mail to colleagues within the company on a workplace-related matter. To do so, test-takers need to:

- comprehend correctly the given inputs (functional knowledge)
- appropriately display knowledge of the use of grammar, syntax and vocabulary (grammatical knowledge)
- organize the message in a coherent and cohesive fashion (textual knowledge)
- reveal ability to make use of appropriate registers, cultural reference and figure of speech according to audience addressed (style awareness)

Prompt Attributes (PA): A writing prompt in the form of an e-mail is provided (see the sample item below) that explains why the test-taker is writing. In the prompt, three guidelines, along with required response length, are provided that explains what content points test-takers should include. When you write a prompt, also consider the following:

a) *Visual input.*
   - o It refers to an e-mail and is provided on the question paper/screen.
   - o This e-mail should look like a real e-mail that includes subject, date, sender's and receiver's names on the top.
   - o The length of the e-mail should be between 45 to 70 words.

b) *Degree of formality.* The purpose of the e-mail is to exchange short and straightforward information between colleagues so the language of the prompt is simple, colloquial and less formal.

c) *Cognitive skills.* According to Bloom's taxonomy, the cognitive demand required to fulfill this task involves only the lowest level of Bloom's taxonomy, the skill of "knowledge". Test-takers demonstrate this skill by means of making a list of the main events, writing a list of any pieces of info they can remember, describing what is happening, etc. Therefore, the input prompt (i.e. e-mail) does not ask test-takers to summarize, analyze or evaluate but simply to describe, tell, list, name, request and confirm.

d) *Topic.* Suggested topics are made according to the results of needs analysis and expected cognitive demand required:

Topic	Example
Exchanging info	confirm shipment order with clients
Requesting info	request status of manual preparation

e) *Guidelines on content points being covered.* Guidelines that explain three required content points are included.
   - First guideline: to confirm the requested info
   - Second guideline: to list the requested items
   - Third guideline: to make a request

Response Attributes (RA): Test-takers will first read the given inputs (i.e. e-mail) and write a reply, in the form of an e-mail, accordingly.

a) *Length of response.* Test-takers need to write a minimum of 50 words. This requirement is shown on the question paper/screen.
b) *Difficulty level.* Given the requirements noted in the PA above, the difficulty level for this task is expected to be the easiest.
c) *Degree of formality.* Given the varying functions of e-mail in daily communication, the expected response is not restricted to any particular business correspondence format. The response can be either formal or informal.

Sample Item (SI):

You have received an e-mail from your colleague regarding the preparation of entertaining the company founder.

Date: Wed 15 Feb 11:20 CST 2009
From: Emma <emma2008@yourcompany.com>
Subject: Company founder's visit
To: Justin <justin1209@yourcompany.com>

Hi, Justin,

I'm writing to ask if there'll be five people, including you, who will give an office tour to Mr. Smith, the company founder, on the day he visits here. Also, could you please let me know where you plan to take Mr. Smith for lunch so I can make a reservation?

Let me know if you have any other special request.

Thank you,
Emma

Write a reply to
- confirm the number of people who will escort Mr. Smith
- list several restaurants you recommend
- request a shuttle bus service from the restaurant to the office

On the next page, please write a minimum of 50 words. You have 12 minutes for this task.

2. Item type 2-Write a memo

General Description (GD): Writing a business memo is found to be one of the most frequent writing tasks in the workplace setting. A business memo helps members of an organization to communicate with each other in an efficient and effective way. Therefore, this portion of the test requires test-takers to demonstrate their ability to write a business memo. To do so, test-takers need to:

- comprehend correctly the given inputs (functional knowledge)
- appropriately display knowledge of the use of grammar, syntax and vocabulary (grammatical knowledge)
- organize message in a coherent and cohesive fashion (textual knowledge)
- reveal ability to make use of appropriate registers, cultural information and figures of speech according to the audience addressed (style awareness)

Prompt Attributes (PA): A business memo is written to perform a specific purpose, such as informing readers of specific information, requesting an action, or reacting to a situation/problem. It is written for a specific audience; therefore it is important to connect its purpose to the interests and needs of the audience. When you write a prompt, consider the following:

a) *Visual input.*
It refers to any kind of form used in workplace, such as an application form, evaluation form, leave application etc.

b) *Targeted audience.*
The targeted audience or readers refer to colleagues or upper management. Clients are not the targeted readers as the memo is intended to be used internally.

c) *Cognitive skills.*
According to Bloom's taxonomy, the cognitive demand required to fulfill this task involves higher levels of Bloom's taxonomy, which refer to skills of "synthesis", and "analysis". Test-takers are expected to demonstrate the skills by putting parts together to form a whole, explaining patterns and meaning and gathering information from given data, etc. Therefore, the prompt requires that test-takers analyze, compare, outline, integrate, and organize etc. The highest cognitive demand, including evaluation, critique and discrimination, is not required in this task.

d) *Guidelines on content points being covered.* Guidelines that explain three required content points are included.

- First guideline: to describe the main purpose(s) of the given input (i.e. a form)
- Second guideline: to describe the form format
- Third guideline: to give general instructions for completing the form

Response Attributes (RA): Test-takers will first read the given inputs (i.e. a form), three guidelines for writing a business memo and write a memo accordingly.

a) *Length in response.* Test-takers need to write a minimum of 120 words. This requirement is shown on the question paper/screen.

b) *Difficulty level.* Given the requirements noted in the PA above, the difficulty level for this task is expected to be intermediate, which falls between the other two tasks (i.e. write an e-mail and write a report).

Sample Item (SI):

You are a departmental head. You have been asked to write an announcement memo informing your team members of a new job analysis form. After reviewing the form below, please write the memo to your department.

Your memo should include the following information:
- main purpose(s) of the form
- brief description of its format
- general instructions for completing the form.

Please write a minimum of 120 words. You have 18 minutes for this task.

****

Please complete this form prior to requesting placement of a wanted ad and return it to the department of human resources. This form aids you in the employee selection process.

Job Analysis

Prepared By ____________________ **Date** ________________________
Title __________________________ **Department** ____________________

Job Title:	Reports to:

Education/Experience Required

Goals/Objectives of Position

Knowledge/Skills Required

Physical Requirements

Special Problems/Hazards

3. Item type 3-Write a report

General Description: Reports of various kinds, such as evaluation report, summary report, financial report etc, are found to be a common type of writing task in the workplace setting. This portion of the test thus requires test-takers to demonstrate their ability to write a report to potential or existing clients on a business-related matter. To do so, test-takers need to:

- comprehend correctly and evaluate analytically the graphic information and express in words(functional knowledge)
- appropriately display knowledge of the use of grammar, syntax and vocabulary (grammatical knowledge)
- organize message in a coherent and cohesive fashion (textual knowledge)
- reveal ability to make use of appropriate registers according to audience addressed (sociolinguistic knowledge)

Prompt Attribute: A writing prompt is provided that explains who the intended reader is, namely potential and existing clients, and what the graphic inputs are. In the prompt, three guidelines, along with the required response length, are provided that explains what content points test-takers should include. When you write a prompt, also consider the following:

a)*Visual/Graphic inputs.*

- o The first graph is a pie chart indicating a total sales/profit/cost amount of product of interests. The other is a bar chart where a comparison of sales/profit/cost growth between current year and the previous year(s) is given.
- o The two graphs will be shown on the question paper/screen.
- o Dos and Don'ts: A report given to potential or existing clients normally conveys a positivemessage. Hence, the second graph, suggesting the growth trend, should present a positive instead of a negative growth number in client company's area.

b) *Cognitive skills.* According to Bloom's taxonomy, the cognitive demand required to fulfill this task involves higher levels of Bloom's taxonomy, which refer to skills of "synthesis", "analysis" and "evaluation". Test-takers are expected to demonstrate the skills by explaining patterns and meaning, predicting and drawing conclusions and making recommendations, etc. Therefore, the prompt requires test-takers to analyze, infer, integrate, assess, and justify etc.

c) *Topic.* Suggested topics are made according to the results of needs analysis and expected cognitive skills required:

Topic	Example
Work report	progress report, performance report

d) *Guidelines on content points being covered.* Guidelines that explain three required content points are included.
   - First guideline: to describe each given graph
   - Second guideline: to analyze any noticeable trends
   - Third guideline: to conclude and recommend according to the graphs

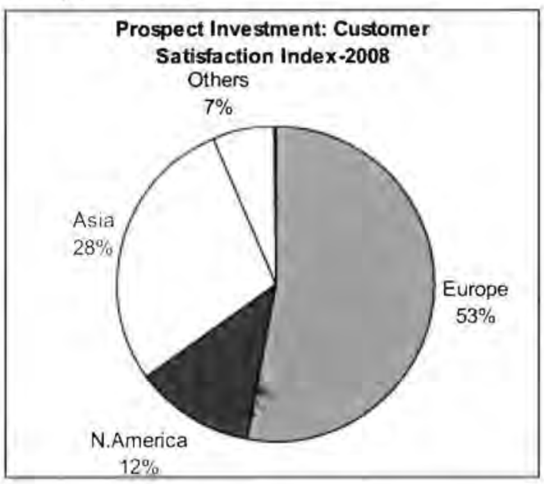
Response Attribute: The test-takers are expected to comprehend the graphic information and express it in words.

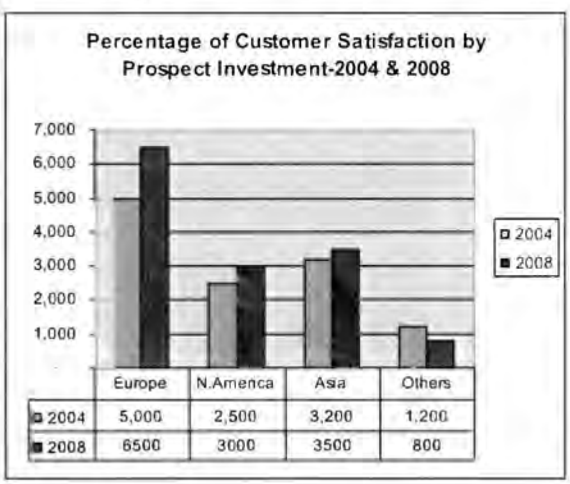
a) Length in response. Test-takers need to write a minimum of 200 words. This requirement is shown on the question paper/screen.

b) Difficulty level. Given the requirements noted above, the difficulty level for this task is expected to be the most difficult.

Sample Item:

You are asked to write a report, to a potential American retailer, analyzing a survey result regarding customer satisfaction in year 2008 and its comparison of year 2004 with year 2008.





	Europe	N.America	Asia	Others
2004	5,000	2,500	3,200	1,200
2008	6500	3000	3500	800

This report should include the following information:

- brief description of each chart
- comments on any noticeable trends
- conclusions and/or recommendations based on the data

On the next page, please write a minimum of 200 words. You have 30 minutes for this task.

PART 2. BTWS RATING SCALE

1. Analytical rating scale

1.1 Assessment criteria

Adapted from the Common European Framework of Reference (CEFR) scale, the BTWS rating scale has been continuously modified according to the outcomes of three activities to better reflect the context of the BTWS: the needs analysis, pilot testing and pilot rating studies. Three assessment criteria were selected in the BTWS rating scale (see Appendix E):

- Task Fulfillment

  This is intended to assess examinees' understanding of given input for each task and ability to perform task requirements accordingly.
- Idea Organization

  This is intended to assess ability to organize writing with the use of organizational patterns, cohesive devices and transition words. Also it assesses ability to develop ideas with supporting details and examples.
- Language Use

  This is intended to assess examinees' grammatical, syntactic and vocabulary knowledge.

1.2. Level of proficiency

Guided by the CEF scale, the BTWS rating scale has three broad divisions: A, B, and C, which can be further divided into six levels: A1, A2, B1, B2, C1 and C2. Expected performance in each division is as follows:

- Level A: examinees exhibit a limited or basic ability to understand task input, to organize test structure and to demonstrate use of grammar and vocabulary.
- Level B: examinees display ability to perform task requirements, to develop ideas appropriately and to use grammar and vocabulary accurately.
- Level C: examinees show ability to perform task requirement efficiently and develop ideas and overall structure of text effectively. They can maintain consistently high degree of grammatical accuracy and use variety of vocabulary.

2. Operational Scoring Method

Two scoring methods, namely analytic scores and an average score that serves as a holistic score, are adopted for rating examinees' essays in the operational scoring system. In the analytic scoring, raters assign three different scores per essay according to the three assessment criteria in the BTWS rating scale. Subsequently, a holistic score is calculated by averaging three analytic scores. Raters do not mark a holistic score independently.

Training guidelines specify each step of rating procedures. See Appendix F, Guidelines for Rater Training, for a suggested training package.

APPENDIX A
Summary of the BTWS task attributes

Task Considerations	**Task 1: Write an Email**	**Task 2: Write a Memo**	**Task 3: Write a Report**
1. Communicative function	Request	Announce	Explain
2. Purpose	Convey objective information	Synthesize given information	Evaluate given information
3. Cognitive demand	Knowledge (lowest)	Analysis and synthesis (intermediate)	Analysis, synthesis and evaluation (highest)
4. Audience	Colleagues	Managers	Clients
5. Prompt input	E-mail (written)	Form (written)	Chart and graph (graphic)
6. Response length	Minimum 50 words	Minimum 120 words	Minimum 200 words
7. Test taking time	12 minutes	18 minutes	30 minutes
8. Expected formality[1]	Less formal --More formal		
9. Expected level of difficulty	Easy --Difficult		

[1] Note: although we present a scale of expected formality for the three task types, in real world use, this attribute may fade. It is possible to have a short but formal e-mail, and likewise, it is possible that a technical report (perhaps due to the norm of communications in a particular business) would be rather informal. Berlitz should monitor this formality scale as the BTWS rolls out, in order to see if formality does align with the tasks as set forth here.

APPENDIX B
Research on writing tests available on the market

1. Reviewed writing tasks

- TOEIC
- BULATS
- TOEFL
- BEC

2. Foci of analysis

- Target level/audience
- Task
- Scoring criteria
- Scoring report
- Link to CEFR
- Other (strength, special observation)

1. TOEIC (www.ets.org)

Target Audience	• People who are preparing to enter the workforce • Individuals who want to measure their speaking and writing skills in English • People who use English in workplace settings such as offices, hotels, hospitals, restaurants, international meetings, conventions and sports events • Managerial, sales and technical employees in international business, industry, commerce and government	
Task	Task type	Task 1: Write a sentence based on a picture (Q1-5) Task 2: Respond to a written request (Q 6-7) Task 3: Write an opinion essay (Q8)
	Directions	Clear and detailed directions • Allowed test-taking time • Information on scoring criteria • Requirements of written response
	Length	60 minutes (total)
	Delivery	Computer
Scoring criteria	Task 1 • Grammar • Relevance of the sentences to the pictures	
	Task 2 • Quality and variety of your sentences • Vocabulary • Organization	
	Task 3 • Whether your opinion is supported with reasons and/or examples • Grammar • Vocabulary	

	• Organization
Score reporting	• Questions 1-5 are rated on a scale of 0-3 • Questions 6-7 are rated on a scale of 0-4 • Question 8 is rated on a scale of 0-5 • Sum of all ratings is converted to a scaled score of 0-200 • 9 proficiency levels are provided • Report sent 3 weeks after testing
CEFR linking	No information on the linking of 9 proficiency levels to CEFR
Strength	• Using multiple writing tasks for diverse skills • Close link between task directions and rating criteria • Detailed descriptors of 9 proficiency levels (including specific writing features and typical flaws) • Provision of guidance and elaboration in task directions

2. Bulats (www.bulats.org)

Target Audience	<General purpose> • Evaluating the language skills of staff within a company • Evaluating the language skills of job applicants • Placing learners on suitable courses for language training • Screening learners who are unsuitable for the training courses provided • Evaluating language training given • Recommending suitable standard examinations for learners • To provide testing support for language audits	
Task	Task type	Part 1: Short Message/Letter (50-60 words) The candidate is given a short text, such as a letter, memo or advert, and guidelines for writing a reply or follow-up message or letter. Part 2: Report or Letter (180-200 words) The candidate writes a short report or letter following brief instructions. For this part, candidates choose a task from two alternatives.
	Directions	• Information about required word limit • No information about assessment criteria • Specific requirements for written response
	Length	45 minutes (total)
	Delivery	Paper (?)

Scoring criteria	Candidates are assessed independently by two trained assessors. They are assessed on: • how accurately and appropriately they use the language (grammar and vocabulary) • how well they organize their ideas • how effectively the piece of writing achieves its purpose
Score reporting	Candidates are given a Test Report which gives an assessment of overall ability as well a description of what they are likely to be able to do in the language in the workplace.
CEFR linking	Candidates receive a Test Report, which gives an assessment of their overall ability (on the ALTE level system of 0 - 5) with finer distinctions of +, - and =. A description of what the candidate is likely to be able to do in the language in the workplace (e.g. 'can deal with clients by responding to simple questions') can be printed on the reverse of the Test Report.
Strength	• Candidates' selection of a topic from two options • Close link to CEFR

3. TOEFL (www.ets.org)

Target audience	• ESL/EFL learners who demonstrate the ability in producing academic English at the university level. • ESL/EFL learners who wish to obtain an admission from undergraduate or graduate college.	
Task characteristics	Task 1: Integrated task	Task 1: 1) *Purpose*: to estimate an ability to understand the main idea and to write an academic style writing by combining sources of two different inputs based on what is read and heard (Reading – listening- writing in this order) 2) *Prompt*: - Reading passage firstly is given on the screen during three minutes with a short direction - Then, a lecture on a topic related to a reading passage is provided - After reading and listening sessions, a question and a short direction how to response appear on the screen. -Length: for reading -18 sentences, and for listening – 23 sentences 3) *Response Format*: an academic writing style -Length for expected responses: 150-225 words
	Task 2: Independent task	Task 2 1) *Purpose*: to estimate an ability to write an opinion on a given topic and to present evidence to support logic of ideas 2) *Prompt:* A simple question with 3 or 4 sentences long are provided that explains why examinees agree or disagree with a given question with a general topic 3) *Response Format*: an academic writing style -Length for expected responses: a minimum of 300 words
	Test delivery	• Computer based
	Time allowed	• Total amount of time: 50 minutes • 20 minutes for Integrated tasks • 30 minutes for Independent tasks
Scoring rubrics	• Score scale : 0 to 5 points (6 levels) • Total of ratings will converts to 0 to 30 score ranged	

	• (Writing score is reflecting 25% for a total score) • Three assessment criteria are used for evaluation; Content (accuracy and completeness), Appropriate use of language, and Sentence structure • Essays are evaluated by four raters
Score report	• Scores on four different skills and a total score are provided • Scores are reported through online and also some feedback on strengths and weakness are provided for future improvement.
Linking to CEFR	• The TOEFL iBT is aligned with Levels A2 to C1 of the CEFR • CEFR TOEFL iBT (Writing section) • A2 = 11 • B1 = 17 • B2 = 21 • C1 = 28
Test preparation	• Sample questions are provided for test preparation • Sample responses for writing are provided • Practice exam can be purchased at online, and examinees can take a practice test and get a instant score. • Notes-taking during reading and listening of the task 1is allowed
Strength	• Multiple writing tasks(Integrated and independent tasks) using multiple inputs(reading and listening) • Provision of clear task directions • Detailed descriptors of 6 level for each tasks • Close link between task directions and rating criteria • Authenticity of two tasks on the academic setting • Close link to CEFR

4. BEC (www.cambridgeesol.org)

Target audience	• Anyone who is preparing for a career in business • Individual learners who wish to obtain a business-related English language qualification	
Task characteristics	BEC components	Reading, Writing, Listening & Speaking
	Proficiency levels	Three different levels offered: • BEC Preliminary • BEC Vantage • BEC Higher.
	Written task type-**BEC Preliminary**	Task 1 -*Purpose*: to produce a piece of communication with a colleague or colleagues within the company on a business-related matter -*Response Format*: to write a note, message, memo, or e-mail. -*Prompt*: o a task prompt provided that explains why examinees are writing and who they are writing to. o three additional prompts also provided that explains what content points examinees should include. o Layout of memos, emails, etc. provided on the question paper. -*Length*: 30-40 words Task 2: -*Purpose*: to produce a piece of correspondence with somebody outside the company (e.g.a customer or supplier) on a business-related matter - *Response Format*: to write a response to a short text, such as a letter or ad - *Prompt:* o a written prompt presented in the form of a short text o four additional prompts also provided that explains what content points examinees should include -*Length*: 60-80 words
	Written task type-**BEC Vantage**	Task 1 -*Purpose*: to produce a piece of communication with a colleague or colleagues within the company on a business-related matter -*Response Format*: to write a note, message, memo, or e-mail. -*Prompt:* o a task prompt provided that explains why examinees are writing and who they are writing to. o three additional prompts provided that explains what content points examinees should include.

		-Length: 40-50 words Task 2: *-Purpose*: to produce a business correspondence with a colleague or colleagues within the company on a business-related matter *-Response Format*: to write a business correspondence, short report or proposal. *-Prompt*: ○ an explanation of the task and one or more texts as input material provided *-Length*: 120-140 words
	Written task type-**BEC Higher**	Task 1: *-Response Format*: to write a short report based on graphic input *-Prompt*: A realistic situation in which examinees have to analyze graphic info and express it in words *-Length*: 120-140 words Task 2: *-Response Format*: examinees choose to write a longer report, a piece of business correspondence, or a proposal *-Prompt:* ○ three task prompts provided that explains why examinees are writing, who they are writing to, and what response format you are writing. ○ three additional prompts provided that explains what content points examinees should include *-Length*: 200-250 words
	Time allowed	• BEC Preliminary: Reading and Writing are combined on one question paper. Length of reading & writing sections: 1 hour 30 minutes • BEC Vantage: 45 min • BEC Higher: 1 hour 10 minutes
	Test delivery	Paper & computer
Scoring rubrics	Not found on the website. Sample writing and examiner comments provided on three different levels.	
Score report & Certificate	• A single, overall grade is awarded, based on the aggregate of scores gained in the four components (Reading, Writing, Listening and Speaking). No Pass/Fail marks for individual papers, so examinees do not need to reach a particular level in any component in order to achieve a Pass in the examination. • Writing section accounts for 25% of the total scores regardless of proficiency level. • Examinees will also receive a Statement of Results which shows your performance in each paper against the scale Exceptional — Good — Borderline — Weak. The report will indicate examinees' relative performance in each skill. If they pass, they will be awarded a certificate from University of Cambridge ESOL Examinations.	
Linking to	• The BEC exams are aligned with Levels B1 to C1 of the CEFR.	

CEFR	• BEC is also linked with the UK Qualifications and Curriculum Authority's National Standards for Literacy, within the National Qualifications Framework (NQF).
Recognition	BEC's website claims that " BEC is officially recognized by more than a thousand educational organizations, employers, ministries, government bodies and professional organizations throughout the world as a suitable qualification for business use"... " Leading international companies such as Sony Ericsson, Shell, Vodafone, Bayer, Coca-Cola and HSBC have all recognized BEC in their offices around the world".
Strength:	• Use multiple tasks appropriate to varying proficiency levels • Reflect authenticity of writing tasks on the workplace • Provide clear task prompts • Provide internationally recognized standards to employers about skills relevant to using English in the workplace. • Close link to CEFR

APPENDIX C
Findings of the needs analysis

Selected results of survey questionnaire

Table 1. Types of writing activities

Items	Never	Rarely	Sometimes	Always	Total
b1 Write reports on a working plan or progress	4.8	26.2	31.0	38.1	100
b2 Write summaries/reviews of written documents	4.8	28.6	23.8	42.9	100
b3 Write technical reports/documents	16.7	21.4	19.0	42.9	100
b4 Write presentation materials	7.1	28.6	23.8	40.5	100
b5 Write short notes/memos/messages	7.1	2.4	9.5	81.0	100
b6 Write meeting notes	9.5	14.3	35.7	41.5	100
b7 Write letters/email to people in your organization	4.8	7.1	11.9	76.2	100
b8 Write letters/email to clients or other organizations	0.0	9.5	14.3	76.2	100

Table 2. Purpose of writing activities

Item	Not relevant at all	Less relevant	Relevant	Very relevant	Total
c1 Write to persuade people	9.5	26.2	42.9	21.4	100
c2 Write to express my own ideas/opinions	0.0	9.5	35.7	54.8	100
c3 Write to demonstrate my knowledge/understanding of a topic	2.4	7.1	31.0	59.5	100
c4 Write to develop an argument	14.3	38.1	35.7	11.9	100
c5 Write to convey objective information	7.1	2.4	23.8	66.7	100
c6 Write to record information for future retrieval	4.8	16.7	33.3	45.2	100

Table 3. Important writing skills

	Not important	Less important	Important	very important	Total
d1 Ability to understand reader's needs, values, and attitude	0.0	4.8	19.0	76.2	100
d2 Ability to organize ideas in a coherent manner	0.0	2.4	9.5	88.1	100
d3 Ability to use a range of vocabulary	0.0	40.5	42.9	16.7	100
d4 Ability to use grammar accurately	2.4	19.0	45.2	33.3	100
d5 Ability to use appropriate style and	0.0	9.5	31.0	59.5	100

tone					
d6 Ability to convey information in a concise manner	0.0	2.4	21.4	76.2	100

Table 4. Writing inputs

Items	Never	Rarely	Sometimes	Always	Missing	Total
e1.email	0.0	0.0	9.5	90.5	0.0	100
e2.graphic	4.8	23.8	42.9	28.6	0.0	100
e3.paper	2.4	16.7	42.9	38.1	0.0	100
e4.audio	16.7	19.0	31.0	33.3	0.0	100
e5.none	42.9	28.6	16.7	2.4	9.5	100

Table 5. Mode of written communication

Item	Never	Rarely	Sometimes	Always	Total
f3_1computer	2.4	2.4	0.0	95.2	100
f3_2handwritten	16.7	50.0	33.3	0.0	100

APPENDIX D

Findings of the pilot test

Part I. Selected findings of the survey questionnaire

Table 6. Satisfaction of the pilot test

Items	Strongly Disagree	Disagree	Agree	Strongly Agree	Miss -ing	Total
b1. Generally I am satisfied with the test.	3.8	3.8	69.2	23.1		100
b2. Three tasks (Tasks 1, 2, and 3) reflect major writing tasks required in a workplace setting.	0	0	46.2	53.8		100
b3. The test seems to successfully measure a test-taker's ability to write in a workplace setting.	0	3.8	57.7	34.6	3.8	100
b4. Test instructions are clear	3.8	38.5	34.6	23.1		100
b5. Total test-taking time (60 minutes) is sufficient	0	19.2	38.5	42.3		100
b6. It was interesting to perform this test	0	11.5	38.5	50.0		100
b7. A computer-delivered format is appropriate for this writing test	0	3.8	38.5	57.7		100
b8. Assessment criteria used in the test are reasonable	0	3.8	65.4	23.1	7.7	100
b9. My reading ability significantly affects my writing performance on this test	3.8	15.4	42.3	38.5		100

Part II. Selected findings of focus group interview

Task Difficulty

With a few exceptions, the task difficulty of four writing types indicated by the interviewees was quite consistent. The majority considered email the easiest type; two of the interviewees weighed email and memo equally easy. The more difficult types mostly referred to technical and work progress report.

Table 7. Rank-ordering of task difficulty

Interviewee	The most difficult		The easiest
1	- - -	E-mail	Memo
2	Memo	Technical report	E-mail
3.	Technical report & work progress report	- - -	E-mail, Letter & Memo
4	Technical report	Memo	E-mail
5	Technical report &	Memo	E-mail

	Work progress report		
6	?	?	E-mail
7	Email	Technical Report & Work progress report	
8	Progress report	Technical report	E-mail
9	Technical report	Work Progress report	E-mail & Memo

The following are the justifications for ranking by several interviewees. Note that all the quotations in this portion of the report reflect the major points indicated by the interviewees; wording may be rephrased by researchers because the interview was not tape-recorded.

Interviewee 2: E-mail is the easiest 'cause I *write it everyday*. Then it's technical report '*cause I've done a lot and there are certain formats to follow*. I also apply my knowledge into the report. The most difficult one is memo '*cause it needs to be concise*.

Interviewee 7: . . . e-mail is more difficult than technical report and work progress report *because e-mail requires an ability to write down the main points of what I want to say*.

Interviewee 8: I did not write a memo frequently. *I feel that it is outdated*. Usually e-mails cover all functions. . . Based on my experience, writing e-mails is the most important *because my writing ability is judged by e-mails*. I remembered that when I first e-mailed to a UK client, I got feedback from a manager about my writing.

Interviewee 9: Writing e-mails is the easiest since *it is a daily-based work*. Writing a memo is similar. . . I think that writing a technical report is the most difficult since it *includes collecting data, analyzing data, and reporting the result*. And I need to make audience understand the content. And sometimes I need to avoid using technical terms and make terms simpler for different audiences.

<u>Writing Skills</u>

Skills that are necessary to effectively write different task types are further explored and indicated below:

E-mail/memo/letter
Conciseness (x 7)
Grammar (x 3)
Straightforward (x 2)
Accuracy (x 2)
Effectiveness/Responding in time (x 2)
Content (x 1); Clarity(x 1); Politeness(x 1); Appropriateness(x 1); Topic (?)(x 1)

Whether or not writing is affected according to audience addressed is revealed in interviewees' different viewpoints.

Writing is affected when addressing different audience:
Interviewee 8: In terms of audience, *I write differently when my audience is top management and when the audience is my colleagues*. When I write to higher levels, my writing becomes more polite and professional.

Interviewee 9: . . . *I consider culture* when I write e-mails/letters. For example, when I write to an American client, I try to be concise, whereas I try to write long and to be polite when I write to a Japanese client. I learned this in a intercultural communication course.

Writing is not affected according to audience addressed:
Interviewee 1: . . . The way I wrote was pretty much *content-oriented*, regardless of audience I addressed, in the sense that biz writing tended to be formal and writing to colleagues was informal.

Work progress report and a technical report
To describe graphical inputs in details (x 4)
To be concise (x 2)
To convey the main points(x 1); organization(x 1); clarity(x 1)

In terms of audiences who most frequently read the reports, supervisors(x 5) are normally the readers, followed by clients(x 3). As for inputs for the report, various types were found, including market data in the form of either a chart or graph(x 2), previous reports(x 1) and statistical findings(x 1).

It was noticeable that, according to the responses above, conciseness is frequently considered to be essential if a good piece of writing is to be produced. Techniques that help accomplish this goal include the following:

Write short sentences/Preciseness (x 3)
Use graphic charts and tables (x 3)
Be straight to the points/Relevance (x 3)
Edit/Proofread (x 2)
Be organized (x 1)

As far as the ability to use a wide range of vocabulary in writing, the interviewees elaborated on why it is treated as less of a priority in business writing. The reasons are as follows:

There is business jargon or terms that you use frequently (x 4)
The importance lies in whether you can convey what you intended to say (x 4)
Key/simple words are sufficient enough to help understanding (x 2)

Furthermore, to prepare potential employees for the future business market, the most common writing skill that interviewees considered essential is to demonstrate analytical skills, such as logic development and organization skills. Others are as follows:

Equip with analytical skills (logic, organizational skills) (x 4)
Familiarize themselves with business jargon (x 1)
Write concisely (x 1)
Know how to get their meaning across (x 1)
Understand writing conventions of different writing tasks (x 1)

In addition, unlike other questions where a variety of responses arise, all interviewees regard the skill of presenting arguments as unnecessary and should be avoided in business writing. If arguments are inevitable, face-to-face communication is more desirable as indicated by most interviewees.

Input

As writing may be carried out according to visual and/or audio input, the interviewees specified different types of input in the forms of graph, paper and audio, respectively:

Graphic input, including number, graph, chart and table and excel data table (x 8)
Paper input, including report, meeting notes, technical specification for vocabulary, Google search and statistics. (x 5)
Audio input, including discussions, phone calls (x 4)

Given the variety of input normally involved in the writing, all the interviewees, when being asked to what extent the reading and listening ability is indispensable to their writing performance, indicated that reading/listening and writing skills go hand in hand. In other words, "a good reader is a good writer", said by Interviewee 9.

Writing process

According to the interviewees, writing process in terms of whether or not the writing begins with a draft is dependent upon several factors: formality(x 2), importance (x 2), readers (1) and time constraints (x 1).

Formality refers to the types of writing task. "If I write an e-mail, I simply don't begin with drafts; otherwise, I'll be more cautious and edit the writing before it's been sent out", said by Interviewee 4.

The importance of the writing is another factor that prompts several interviewees to write a draft and get feedback from colleagues.

The reader is also a key factor. For example, Interviewee 6 indicated that the first thing that comes to his mind when writing is to think of who is going to read his writing. Then he will decide whether he should begin with a draft.

Finally, time constraints were said to affect the writing process. That is to say, if it is routine writing, then a draft may not be necessary. However, if the writing is under time pressure, usually the writing is done without a draft.

Test-taking experience

The interviewees expressed their perceptions on the evaluation criteria that should be employed in an ideal test of business writing skills. These include:

Grammar (x 3)
Conciseness (x 3)
Organization (flow, logical structure) (x 2)
Main ideas (x 2)
Writing convention/Formality (x 2)
Straightforwardness (x 1)
Address according to audience (x 1)
Rhetorical style (x 1)
Knowledge of content (x 1)
Business language (x 1)

As far as the difference between business writing and other types of writing, the interviewees referred to 'Conciseness' as a salient feature, followed by 'neutral tone' in the sense of the flexibility to rephrase negative information, such as a rejection letter to a candidate, into a more positive tone. The distinct features of business writing are as follows:

Conciseness (x 4)

Neutral tone (x 2)
Content (x 1)
Logical organization (x 1)
Writing conventions (x 1)
Simply words (x 1)
Use graphic info (x 1)
Clarity (x 1)
Professional (x 1)
Not much different (x 1)

Conclusion. Implications for defining constructs, modifying the CEFR rating criteria, describing prompts and justifying integrated skills.

Writing skills identified	Constructs to be measured	Modifications of the CEFR rating criteria	Description of prompts	Expected response
Conciseness	X	X		X
Grammar	X			X
Full descriptions of graphical input		X		X
Conveying intended meaning	X	X	X*	X
Organization(flow, logical structure)	X			X
Writing conventions				
Neutral tone		X		X

APPENDIX E

BTWS analytical scale

Level	Task Fulfillment *Essays at this level…*	Idea Organization *Essays at this level…*	Language Use *Essays at this level…*
	Assess examinees' • Understanding of given input • Fulfillment of content requirements • Style/audience awareness	Assess examinees' • Ability to organize text using cohesive devices and to plan text following conventional structure of essay (i.e. introduction, body and conclusion) • Ability to group and present ideas with supporting details/examples	Assess examinees' • Grammatical and syntactic knowledge • Vocabulary use • Correctness in spelling
C2	• Display complete understanding of given input prompts • Fully perform all content requirements • Display competent style/audience awareness	• Plan text fully using conventional structure of essay that includes introduction, body and conclusion and create organized and coherent text consistently using a range of cohesive devices and transition words • Develop a seamless flow of ideas with relevant supporting details and/or examples that show depth and insight.	• Maintain consistent grammatical control and show great flexibility formulating sentence structures • Use vocabulary precisely to give emphasis, to differentiate and to eliminate ambiguity. • Show a consistent degree of competency in spelling
C1	• Display good understanding of given input prompts • Effectively perform almost all content requirements • Display effective style/audience awareness	• Organize text following conventional structure of essay that includes introduction, body and conclusion and effectively show controlled use of cohesive devices and transition words • Develop a smooth flow of ideas with relevant supporting details and/or examples, though a few points may not be fully elaborated.	• Effectively maintain a high degree of grammatical accuracy and have a good command of formulating sentence structures; errors are rare and difficult to spot • Use vocabulary flexibly on professional and general topics • Show correctness in spelling; errors are rare
B2	• Display sufficient understanding of given input prompts	• Plan text appropriately that shows a clear structure of essay and use a limited number of cohesive devices to link sentences into	• Show a relatively high degree of grammatical control and clearly formulates sentence structures;

	• Appropriately perform content requirements even though some points are missing • Display appropriate style/audience awareness even though minor discrepancies are observed	clear, coherent text, though there may be some 'jumpiness' in a long construction; • Appropriately develop ideas with supporting details and/examples, though some points may be less clearly elaborated.	errors do not cause misunderstanding • Have sufficient vocabulary to express viewpoints • Can spell most words correctly
B1	• Display reasonably acceptable understanding of the given input prompt • Attempt to fulfill content requirements, but they are partially performed. • Display reasonably acceptable style/audience awareness, but some inappropriate uses were observed	• Structure text reasonably accurately and can link a series of shorter, discrete simple elements • Display ideas with limited development in depth and thoroughness.	• Use grammar and frequently used sentence patterns reasonably accurately, though readers may be obliged to figure out the meaning after several iterative readings. • Have enough vocabulary to convey meanings on general topics • Can spell correctly the majority of words
A2	• Display basic understanding of the given input prompts • Fulfill content requirements only to a limited extent. • Display basic style/audience awareness; the awareness is either inappropriate or inconsistent	• Have some basic control of structuring text and link groups of words with basic transition words like "and", "but" and "because". • Present ideas that are understandable through reader's efforts but not expanded.	• Use basic grammar and simple sentence structures correctly; mistakes are noticeable without reader's effort • Use basic vocabulary to communicate limited information • Can spell frequently used words correctly only
A1	• Display limited understanding of the given input prompts • Content requirements are rarely fulfilled. • Display limited style/audience awareness; difficult to identify style/audience awareness	• Show limited control of structuring text and link words or groups of words with limited linear connectors such as 'and' or 'then'. • Present ideas that are developed in a limited manner.	• Show only limited control of a few simple grammatical structures and sentence patterns in a memorized repertoire • Have a very limited repertoire of words • Show difficulties with spelling

APPENDIX F
Suggested Guidelines for Rater Training

Step 1. The Structure of the BTWS: What is the BTWS? (See the Specs: "introduction")

2. Three Different Tasks (Review from the Specs, Part 1, section 1, 2, and 3 for each task type; perhaps have the new raters practice each type)

  3.1 Task 1: Write an e-mail
  3.2 Task 2: Write a memo
  3.3 Task 3: Write a report

3. Assessment Criteria (See the Specs, Part 2, section 1.1; here, raters should practice with actual sample essays)

  5.1 Task Fulfillment
  5.2 Idea Organization
  5.3 Language Use

4. The Rating scale (See the Specs Part 2 and Appendix E; please acquaint the raters with the CEFR in general and with each level)

  4.1 CEFR Scale
  4.2 A1 Level
  4.3 A2 Level
  4.4 B1 Level
  4.5 B2 Level
  4.6 C1 Level
  4.7 C2 Level

5. Prototype Essay for each proficiency level and rating practices
Prototype essays can be used in the training workshop in order to enhance raters' understanding of making scoring judgment. Based on the consensus between stakeholders (i.e. a trainer and raters), more than two representative essays of varying proficiency levels can be selected within each task and can be used in the lecture and discussion session in the training workshop. Prototype essays can be obtained from the Pilot Testing essay bank, and as the BTWS rolls out, from operational test use.

6. Scoring the Essays
This section provides general guidelines on the entire system for methods of scoring examinees' essays. It includes the rating procedures, rating principles, scoring methods, a score report, and rater pairing system.

6.1 Operational Rating Procedure
Examinees' essays should probably be saved on the Berlitz server and will be assigned to individual raters. Individual raters will access folders on the Berlitz server by using the assigned password. The folder for individual raters contains a rating packet such as a particular test form (set), examinees' essays, rating scale descriptor, and a rating sheet for

a score report. The rating results will be automatically submitted to the rating file directly or individual raters will use e-mail to submit the results.

6.2 Rating Principles

6.2.1 Rating Procedure

Before rating, raters should familiarize themselves with the rating scale descriptors for each assessment criteria, each proficiency level, and a particular set of test prompts. The following guidelines describe each scoring step in details.

Step 1: Familiarize yourself with the BTWS rating scale

Three assessment criteria are designed for the BTWS rating scale: Task fulfillment, Idea Organization and Language use. Each assessment criterion has six levels of proficiency, ranging from A1 to C2 (See Appendix E, the Rating Scale).

Step 2: Read writing prompts

Raters should read the test prompts and be familiarized with each task requirements. Pay attention to the context, audience addressed and specific requirements (e.g., word limit and required content for each task).

Step 3: Read examinees' essays on a taskwise basis (not essay-by-essay)

Read task 1 of all essays, followed by task 2 and task 3, respectively. Be sure to familiarize yourself with the specific task requirements before rating.

Step 4: Assign analytic scores

Score task 1 of all essays, followed by task 2 and task 3, respectively. When scoring, assign scores according to the three assessment criteria. Mark the scores on the official rating sheet provided. If you have difficulty in determining scores, look at the rating scale again and make the best judgment you can. Make a note of reasons that result in difficulty in determining scores. The follow-up rater training session will allow you to discuss issues that you encounter when assigning scores.

6.2.2 The Scoring Method and System

Raters should assign three analytic scores across three tasks, and the holistic score is calculated by averaging three analytic scores. The result will be used in a score report with descriptions of each proficiency level for examinees.

6.2.3 Rater Pairings

Berlitz informs and assigns a model for rater pairings via individual e-mails. Two raters participate in the first rating and the third rater's participation is necessary when a discrepancy of one point occurs between scores assigned by the first two raters.

6.3 Institutional rating policy (The Dos and DON'Ts of rating)

- Before starting rating, familiarize yourselves with the overall scoring procedures.
- Fully understand the test prompts and directions before scoring.
- Make scoring judgment according to the rating scale descriptors.

- During rating, the degree of formality of examinees' responses is not the critical rating point.
- Try to find out the advantages and disadvantages of each essay. It is helpful to determine the broad divisions first (A, B, and C) and then the six proficiency levels.
- Do not compare essays during rating, and use criteria that the rating scale provides.
- Do not use your internal standards(e.g. your belief in what constitutes a good writing) that are not included in the rating scale If you have problems in scoring, review the training manual and prototype essays for each level.

7. Other issues

This section provides general information about the rater workshop program such as the role of a trainer, workshop schedule, and continued education program.

7.1 Role of the Trainer

The primary role of the trainer, as a facilitator, is to train new and experienced raters by providing practice sessions during the workshop. Also, the trainer encourages raters to conduct their task and gives advice to solve their problems in scoring. Additionally, the trainer plays a role to deliver raters' messages/concerns to the testing agency.

7.2 Training Workshop Administration

7.2.1 General information for workshop

Berlitz provides a regular training workshop in order to train newly recruited raters. The primary audience is new raters, and the training workshop lasts for four days. After participating in all the workshop activities, candidates will become official raters.

7.2.2 Continued Education

Berlitz provides a continued education program for experienced raters in the form of on-site and on-line rating program. The BTWS trainers or official BTWS raters can request the continued education program. There are two purposes of the continued education: refreshment before rating, or extensive discussions for the problem-solving. The continued education is a useful program for experienced raters to review the rating principles and retrain their skills before the rating. Furthermore, if raters have problems in scoring, then it can be a place to discuss their concerns and share the solutions with each other.

EXHIBIT E

Berlitz Testing

Berlitz Test of Writing Skills
SPANISH

Tarea 1 – Escriba un mensaje de correo electrónico

Usted ha recibido un mensaje via correo electrónico de su colega con respecto a una reservación de una sala para una reunión.

Fecha: Viernes 14/02/2009 15:20
De: Fabio <fabio2008@suempresa.com>
Asunto: Reservación de una sala de conferencia
Para: Jaime <jamie1209@suempresa.com>
Hola Jaime: En referencia a su solicitud, le escribo para informarle que el cuarto de reuniones estará libre toda la mañana del día lunes 9 de Marzo. Le adjunto el formulario de reservación del cuarto de reuniones. Si desea hacer la reservación por favor complete el formulario y envíemelo. También añada una lista de todo el equipo que necesitará para el cuarto de reuniones. Le saluda atentamente, Fabio

Por favor escriba su respuesta en la siguiente página:

- Confirme la reservación.
- Escriba una lista de solicitud de los varios tipos de equipo que necesitará para la reunión.
- Especifique que usted quiere un escritorio de recepcionistas fuera de la sala de conferencia.

Usted debe escribir por lo menos 50 palabras y tiene 12 minutos para llevar a cabo esta tarea.

Berlitz Testing

Fecha: Lunes 17/02/2009 09:40
De: Jaime <jaime1209@suempresa.com>
Asunto: re: Reserva del cuarto de reunión
Para: Fabio <fabio2008@suempresa.com>
(Escriba su correo electrónico en esta página.)

Tarea 2 – Escriba un memorándum

Usted es el gerente del departamento. Le han pedido que escriba un memorándum para anunciarles a los miembros de su equipo acerca del nuevo "Formulario para solicitar ausencia". Después de leer el formulario solicitando la ausencia, por favor escriba un memorándum para su departmento.

Su memorándum tendrá que incluir la siguiente información:

- El propósito principal del formulario.
- Una descripción breve de su formato.
- Instrucciones/indicaciones generales para llenar el formulario

Por favor escriba por lo menos 120 palabras. Usted tiene 18 minutos para terminar esta tarea.

Berlitz Testing

Solicitud de Ausencia

Por favor llene este formulario para la solicitud de ausencia antes de salir de vacaciones o por cualquier otro motivo de ausencia de corto plazo. Usted no estará autorizado para ausentarse del trabajo hasta que reciba un correo electrónico de aprobación.

Sección 1: Información del empleado		
Nombre:	Apellido:	
Puesto de trabajo:	Departamento:	
Teléfono:	Correo electrónico:	
Sección 2: Información detallada de la ausencia		
Clase de ausencia solicitada		
Enfermedad ☐	Vacaciones ☐	Fuerzas armadas ☐
Actuar como jurado ☐	Maternidad/Paternidad ☐	Otro ☐
Fechas de ausencia Desde ____________________ Hasta ____________________		
Razón de ausencia:		
Firma del empleado:	Fecha / /	
Sección 3: Aprobación del supervisor del departamento		
Aprobado ☐	No Aprobado ☐	
Firma:	Fecha / /	
Sección 4: Aprobación del gerente de la división		
Aprobado ☐	No Aprobado ☐	
Firma:	Fecha / /	

- Un certificado médico es requerido después de tres días consecutivos de no asistir al trabajo si su ausencia es por razones de enfermedad.
- Cuando deba cumplir como miembro del jurado o deba ausentarse para enrolarse en el ejército, por favor entregue copias de los documentos correspondientes.
- Usted está obligado a dejar el número de un contacto en caso de emergencia si usted va a viajar.

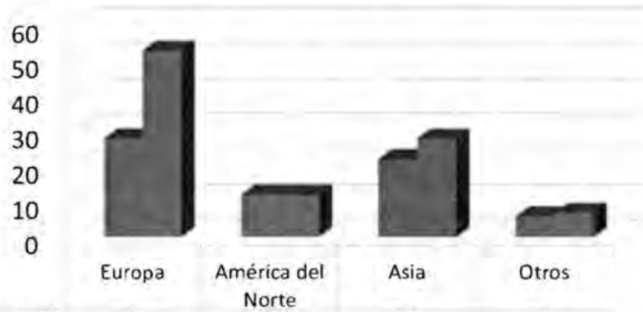
(Escriba su memorándum debajo)

Tarea 3 – Escriba un informe

Usted trabaja para una empresa de ropa. Le han pedido que escriba un informe, para potenciales clientes minoristas europeos, en el que se analiza el crecimiento de las ventas de la empresa. El siguiente gráfico muestra el total de ventas del año pasado y la comparación de ventas de los años 2004 y 2008.

Gráfico 1
Ropa P&P
2009 Monto Total de Ventas por Regiones

Gráfico 2
Ropa P&P – Monto de Ventas
Comparación 2004 v.s. 2009 (Unidad: Millones)



	Europa	América del Norte	Asia	Otros
2004	28	12	22	6
2009	53	12	28	7

Su informe debe incluir la siguiente información:

- Escriba una breve descripción de cada gráfico
- Escriba comentarios referentes a tendencias que usted observe en los gráficos.
- Saque conclusiones y haga recomendaciones para clientes minoristas potenciales.

En la página siguiente, por favor escriba por lo menos 200 palabras. Usted tiene 30 minutos para terminar esta tarea.

(Escriba su informe debajo.)

EXHIBIT F

Berlitz Test of Writing Skills Validation Report

For

Berlitz International Inc.

Prepared by

The University of Illinois at Urbana-Champaign (UIUC)
Foreign Language Assessment Group (FLAG)

Principal Investigator: Prof. Fred Davidson (Linguistics)
Jiyoung Kim (Center for Teaching Excellence)
Tammy Hsu (Educational Psychology)
So-Young Jang (Educational Psychology)

June 10, 2009

TABLE OF CONTENTS

LIST OF TABLES

1.0 INTRODUCTION

The Berlitz Test of Writing Skills (BTWS) was developed through the collaborative efforts of Berlitz International Inc. and the University of Illinois at Urbana-Champaign. The test is designed to assess English writing skills required in a workplace setting. The development of the writing test was guided by a needs analysis and pilot testing. Three forms of the test have been developed. The writing test is intended to be used by organizations, international corporations and government agencies throughout the world in their recruitment, staff training and development, and benchmarking processes.

2.0 OVERVIEW OF THE BERLITZ TEST OF WRITING SKILLS

2.1 Descriptions of the tasks

The BTWS includes three tasks that were based on the outcomes of three activities: research on writing tests available on the market, a needs analysis study and a pilot test. Effort has been made to select task types that best reflect the 'situational authenticity' (Bachman & Palmer, 1996), in the sense that the representative writing tasks in the real business setting are properly selected. The three chosen tasks were: write an email, a memo and a report respectively. In 'Write an email', test-takers are asked to write an email reply according to a colleague's email message. In 'Write a memo', test-takers need to write a short memo announcement to a wider audience in the company with regard to a newly launched policy. In 'Write a report', test-takers are expected to demonstrate their ability to write a professional evaluation report based on given graphs to potential or existing clients. Given a variety of factors that influence the level of difficulty in each task, the expected level of difficulty for 'Write an email' is the easiest while 'Write a report' is the most difficult. See the BTWS specification for description of tasks in Appendix A.

2.2 Description of the scoring system

Adapted from the Common European Framework of Reference, the scoring system of the BTWS was developed on the basis of empirical evidence collected from the pilot rating study where raters from Berlitz International, Inc. and University of Illinois at Urbana-Champaign were involved. Essay samples collected in the pilot test were rated by the raters and their comments were documented in the reflection log, both of which contributed to evaluation of the

BTWS rating procedures as well as changes of descriptor wording at each level of proficiency to better reflect the context of the BTWS.

3.0 VALIDATION CLAIMS

This report presents validity evidence according to argument-based approach to test validation (Weir, 2005; Kane, 2006). The BTWS claims four major validity arguments:

3.1 BTWS involves a dynamic validation process.

3.2 BTWS is developed based on a strong theoretical framework of *Language for Specific Purposes (LSP)*.

3.3 BTWS achieves a high degree of situational authenticity.

3.4. BTWS is assessed with a valid scoring system.

Messick (1989) defines validity as "an evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores or other modes of assessment" (p. 1). The developers of the BTWS actively collected both theoretical rationales and empirical evidence through diverse methods and data. The sources of methods and data used include document reviews, a needs analysis, pilot testing and pilot rating. In the needs analysis, pilot testing and pilot rating, both qualitative and quantitative data were collected and analyzed in the framework of a mixed methods research approach to increase defensibility and accountability of data. See Appendix B (Validity Arguments for the BTWS) for the validation argument structure.

4.0 FINDINGS

4.1. Claim 1- BTWS involves a dynamic validation process

The first validity claim states that the BTWS involves a dynamic validation processes, which is influenced by current trends of validation strategies and approaches that have been significantly influenced by post-Messick view (1989) that portrays validity as a unitary but multifaceted concept, and this perspective has now been widely adopted by language testers. Validity is seen as a process of actively collecting evidence or arguments for a particular test-score interpretation and use, as opposed to being determined predominantly through correlation coefficients, as practiced in the Pre-Messick period.

The four validation claims made in this report are based on *Kane's argument-based* approach (Kane, 2006). This approach to test validation has become more pronounced recently and can be found increasingly in modern validation work (Bachman, 2005; Fulcher & Davidson, 2007; Toulmin, 2003). As such this report provides the logic and a set of procedures for investigating and supporting claims about the use of this test and utilizes a systematic approach that draws valid inferences from a wide range of data. In addition, this report follows mixed methods techniques, namely combining qualitative and quantitative approaches, in an effort to yield richer and better supported interpretations and insights into the proposed validity claims.

The varying levels of inferences that form the basis of the argument for the BTWS validation study include *a priori* validity evidence (Weir, 2005). As the name suggests, a priori validity evidence is collected before the test event and will be further elaborated below in claim 3, examining the authenticity of the BTWS and claim 4, investigating the development of scoring system. Those two claims reflect *construct validity* that stresses the relationship between the BTWS and traits that it intends to measure. The detailed processes of and approaches to evidence collection will be further illustrated in claim 3 and 4.

Equally important is *content validity*, another type of priori validity evidence. The content validity is evidence that includes the documentation of how a test is developed, the various decisions made (e.g. construct, expected prompt and response, test procedures, etc.) during the design process and the reasons for those decisions. This documentation is best explained in a test blueprint or "test specification" (Davidson & Lynch, 2002). The test specification evolves and becomes more fixed to respond to all the information arising from test specification construction stage, pilot test and other studies. It therefore not only forms the basis for test construction and administration but also contributes evidence to address validity. An increasing number of testing scholars express a similar view that every record of test design and development decision contributes to the validity argument (Miselvy, Steinberg & Almond, 2002; 2003).

The BTWS specification has undergone a number of changes, which are informed and shaped by the results of needs analysis study, pilot test and regular discussions between specification writers, namely the Foreign Language Assessment Group, with Berlitz Testing development division. It therefore led to a defensible, theoretically solid and workplace- and market-oriented product, which combines many different perspectives based on theory and practice. See Appendix A for the specification.

4.2 Claim 2- BTWS is developed based on a strong theoretical framework of *Language for Specific Purposes (LSP)*

The second validity claim is that the BTWS is developed based on a strong theoretical framework for Language for Specific Purposes (LSP). LSP is a "branch of language testing in which the test content and test methods are driven from an analysis of a specific language use situation…" (Douglas, 2000, p. 1). Since the BTWS has a specific purpose, that is, assessing second language writing ability in a business setting, the LSP framework is considered appropriate for the BTWS development. The key aspect of LSP is that test tasks and content are "authentically representative" of Target Language Use (TLU) situations (Douglas, 2000). Bachman and Palmer (1996) defines TLU domain as "a set of specific language use tasks that the test taker is likely to encounter outside of the test itself, and to which we want our inferences about language ability to generalize" (p. 44).

Bachman (1990 proposed two aspects of authenticity. The first aspect is situational authenticity that concerns the extent to which a test task simulates real-life texts or tasks. The second aspect is interactional authenticity that concerns the extent to which the test task interacts with the test-takers' specific purpose language ability. The BTWS attempted to ensure both aspects of authenticity. As Douglas (2000) notes, situational authenticity can be demonstrated by showing an explicit relationship between test task characteristics and the task features in TLU situations. Douglas suggests several techniques for analyzing TLU situations, including grounded ethnography, context-based research, and subject specialist informant procedures. The BTWS adopted a needs analysis as a main technique to analyze TLU situations. First, literature on business writing and existing business writing tests were reviewed to help test developers understand features of writing tasks in a business setting. A needs analysis survey was developed based on the literature review and administered to a specific target audience who had at least one-year work experience in a workplace setting where English was used as a medium for written communication. A survey was conducted followed by focus group interviews, in an attempt to further elaborate and expand the findings of the survey. Both the survey and interviews helped the test development team obtain useful commentary from participants in the TLU situation.

The BTWS used pilot testing to gather evidence for interactional authenticity. As Douglas points out, "no matter how careful test developers may be in making the transition from

TLU task characteristics to test-tasks, it is not until the tasks are piloted, tried out on live test-takers, that it can be seen how well they work…" (pp. 246-247). A pilot test was constructed guided by the findings of the needs analysis and administered to survey participants. Test-takers' performance on the pilot test revealed how test-takers interact with the given test tasks and also the likelihood of obtaining the expected responses. In addition, the feedback from the test-takers helped evaluate several aspects of the test, such as appropriateness of test content, the difficulty level of tasks, the clarity of test instructions, and the appropriateness of test-taking time. The results of the pilot test were used for further test revisions and rating scale development. In the following section, the findings from the needs analysis and pilot testing will be discussed in detail.

4.3. Claim 3: BTWS achieves a high degree of situational authenticity

4.3.1.Overview

The purpose of the needs analysis study was to seek information about the extent and types of written communication normally involved in the workplace setting. Participants in the needs analysis study were the international graduate students currently enrolled in the College of Business at University of Illinois at Urbana-Champaign in fall 2008. The selection criterion required participants to have at least one year work experience in the business setting where English is used as a medium for written communication. The needs analysis study included two phases: the first phase was the survey study followed by the second phase, the focus group interview. In the first phase, a total of 42 participants signed up for the survey session. Issues investigated in the survey included: 1) types of writing activities, 2) purpose of writing activities, 3) important writing skills, 4) types of input, 5) frequency and mode of written communication, and 6) themes or topics of written correspondence. A four-point Likert scale was used in the survey questionnaire, along with several open-ended questions in which participants were asked to list themes/topics which they had to address in different writing tasks (i.e., letter, memo, report, e-mail, or others).

In the second phase, focus group interviews in a semi-structured form were conducted. An interview protocol was prepared and emerging probing questions were given according to participants' responses. Nine participants were selected to represent varying backgounds of nationality, specialization and varying length of work experience. The purpose of the interivews was to expand and elaborate on the findings from the needs analysis survey. Since information on task type, writing skills, purpose, and input type were considered most important for test construction, the interviews focused on issues as noted above. Additionally, issues of writing process and test-taking experience were discussed in the interviews.

4.3.2. Major findings of the needs analysis survey and interview

4.3.2.1 Types of writing tasks

Findings of the survey. Questions about types of writing were expected to provide useful information on the types of writing tasks that the BTWS should adopt. The participants were asked to indicate the extent to which they engage in the listed activities. As seen in Table 1, the two most frequent writing activities were writing short notes/memos/messages and writing letters/emails. 81% of the participants responded that they always wrote short notes/memos/messages. 76.2% of the participants answered that they always wrote letters/email to either people in their organization or outside clients. Writing summaries/reviews of written documents and writing technical reports/documents were also found as common writing activities. Over 60% of the participants responded that they either sometimes or always engaged in these writing activities.

Table 1.
Types of writing activities

Items	Never	Rarely	Sometimes	Always
b1 Write reports on a working plan or progress	4.8	26.2	31.0	38.1
b2 Write summaries/reviews of written documents	4.8	28.6	23.8	42.9
b3 Write technical reports/documents	16.7	21.4	19.0	42.9
b4 Write presentation materials	7.1	28.6	23.8	40.5
b5 Write short notes/memos/messages	7.1	2.4	9.5	81.0
b6 Write meeting notes	9.5	14.3	35.7	41.5
b7 Write letters/email to people in your organization	4.8	7.1	11.9	76.2
b8 Write letters/email to clients or other organizations	0.0	9.5	14.3	76.2

Findings of the open-ended questions. The open-ended questions helped to gain in-depth understanding of specific themes or topics for each writing task, namely letter, memo, report and email respectively. It was found the topics on writing a letter are most likely associated to building rapport with existing customers (e.g. a thank you note to customers) and to discuss or report the status of work (e.g. write to clients regarding issues of budget). Others write letters for the purpose of requesting and proposing, such as to request an appointment or to write a resignation letter.

Respondents also indicated varying types of topics on writing memos. The most common type was the meeting notes, followed by writing a memo to request for leave. Others used memos to function as announcements and to send requests for appointments. Next, findings from writing reports suggested that the majority was related to periodic work or status progress report. Also commonly written was the evaluation report, followed by project summary report, technical document and training document. Lastly, the richness of the use of email suggested a wide variety of topics used in the workplace settings. The majority of respondents expressed that email was used to exchange and request information. It was also common to use email to respond to clients' requests or needs and to solve issues raised by colleagues or customers.

Findings of the focus group interviews. In addition to the results from the survey questionnaires, the follow-up interviews further contributed to our understanding of the varying levels of difficulty of the four task types, as noted above. With a few exceptions, the task difficulty of the four writing tasks indicated by the interviewees was quite consistent. The majority considered email the easiest type; two of the interviewees weighed email and memo equally easy. Most regarded technical and work progress reports as the more difficult types.

The following were the justifications for ranking by two of the interviewees, which particularly showed the contrasting views of the level of difficulty of writing emails.

Interviewee 2: Email is the easiest 'cause I *write it everyday*. Then it's technical report '*cause I've done a lot and there are certain formats to follow*. I also apply my knowledge into the report. The most difficult one is memo '*cause it needs to be concise*.

Interviewee 7: . . . e-mail is more difficult than technical report and work progress report *because email requires an ability to write down the main points of what I want to say*.

4.3.2.2 Important writing skills

Findings of the survey. Questions about important writing skills were expected to provide useful information on deciding constructs to be assessed in the BTWS. Participants were asked to indicate the extent to which the listed abilities are important for successful business writing. As seen in Table 2, three skills were found to be very important: 1) Ability to organize ideas in a coherent manner (88.1%), 2) Ability to understand reader's needs, values, and attitude (76.2%), and 3) Ability to convey information in a concise manner (76.2%). Besides the three skills, the ability to use grammar accurately and appropriate style and tone were found important as well. 91.5% of the participants responded that using appropriate style and tone is either important or very important. And 78.5% of participants answered that using grammar accurately is either important or very important. 40% of the participants answered that the ability to use a range of vocabulary is less important, whereas 59.6% responded that the skill is either important or very important.

Table 2
Important writing skills

	Not important	Less important	Important	Very important
d1 Ability to understand reader's needs, values, and attitude	0.0	4.8	19.0	76.2
d2 Ability to organize ideas in a coherent manner	0.0	2.4	9.5	88.1
d3 Ability to use a range of vocabulary	0.0	40.5	42.9	16.7
d4 Ability to use grammar accurately	2.4	19.0	45.2	33.3
d5 Ability to use appropriate style and tone	0.0	9.5	31.0	59.5
d6 Ability to convey information in a concise manner	0.0	2.4	21.4	76.2

Findings of the focus group interviews. The findings from the follow-up interviews indicated somewhat different yet insightful perspectives on skills necessary to prepare potential employees for the future business market. The most common writing skill that interviewees considered essential was to equip oneself with analytical skills, such as logic development and organization skills. Other features were indicated as follows:

Familiarize themselves with business jargon (x 1[1])
Write concisely (x 1)
Know how to get their meaning across (x 1)

[1] (x 1) indicates one out of nine interviewees felt this way

Understand writing conventions of different writing tasks (x 1)

Furthermore, as shown in the survey results that the ability to use a range of vocabulary was less important, as indicated by 40% of the respondents, it led the researchers to further explore interviewees' perceptions on the role of vocabulary in business writing in the interview sessions. It was found that to use a wide range of vocabulary is neither a prioritized concern nor essential skill. Interviewees pointed out two major reasons to justify their perspectives: one was that business writing mostly uses business terms or jargon instead of fancy words that may be used in creative writing; the other was the view that priority lies in whether one can convey what he/she intended to say. Hence, two of the interviewees expressed that simple or key words are sufficient enough to help understanding of messages.

As for the important skills essential to good writing, the findings from the survey questionnaires were quite consistent with those of the follow-up interviews. All interviewees believed conciseness is the most important skill regardless of the writing task. The interviewees further pointed out techniques that helped accomplish this goal, which included the following:

Write short sentences (i.e. preciseness) (x 3)
Use graphic charts and tables (x 3)
Be straight to the point (i.e. relevance) (x 3)
Edit/Proofread after writing (x 2)
Be organized (x 1)

As opposed to the survey results that stressed the importance of the ability to understand intended readers' needs and to organize ideas in a coherent fashion, the interviewees pointed out other skills that were considered essential, such as to be straightforward when conveying intended messages; to express accurate information; to be able to respond in a timely manner; and to describe graphs in details if writing a report.

4.3.2.3 Other important information

Purpose of writing activities. The questions on the purposes of writing activities intended to discover major communicative functions of business writing. Participants were asked to indicate the degree to which the listed purposes of writing are relevant to the writing tasks they had performed before. The findings (See Appendix C) showed that 66.7% of the participants agreed that conveying objective information is a very relevant writing purpose. Demonstrating knowledge/understanding of a topic and expressing own ideas/opinions were also found very relevant writing purposes. Compared to other purposes, developing an argument was not considered a very relevant writing purpose in a business setting. Only 11.9% of the respondents agreed that developing an argument is a very relevant writing purpose.

The last survey findings correspond to interviewees' responses in the follow-up interviews where they pointed out the skill of presenting arguments as unnecessary and impolite and therefore should be avoided in the business writing. Nevertheless, if arguments are inevitable, face-to-face communications are preferable as they may ease tension and allow immediate negotiation with clients or colleagues.

Types of input. The questions about types of input were to discover authentic types of input for the BTWS writing tasks. Participants were asked to indicate the frequency that writing was performed responding to the following inputs: 1) Email/letter input 2) Graphic input, 3) Paper input, 4) Audio input, and 5) No input. The findings (See Appendix C) showed that business writing is usually performed based on a certain type of input. Only 2.4% of the participants indicated that writing is always performed with no input. Among the different types of input, it was found that email/letter input is the most frequent. 90.5% of the participants responded that they always wrote responding to email/letter. Paper, graphic, and audio inputs were found to be commonly used.

Consistent with the survey findings above, findings from the follow-up interviews further suggested specific types of inputs in the forms of graph, paper and audio respectively:

Graphic input: including number, graph, chart and table and excel data table (x 8)

Paper input: including report, meeting notes, technical specification for vocabulary, Google search and statistics. (x 5)

Audio input: including discussions, phone calls (x 4)

It should be noted that given the varieties of inputs normally involved in business writing, all the interviewees, when being asked to what extent the reading and listening ability is indispensable to their writing performance, indicated that reading/listening and writing skills go hand in hand. In other words, "a good reader is a good writer", said Interviewee 9.

4.3.3 Pilot testing and feedback

4.3.3.1. Overview

The purpose of the pilot testing was to try out the initial BTWS tasks and to gather validity evidence for further improvement of the test. Among those who had participated in the needs analysis survey, 28 students voluntarily participated in the pilot testing. Right after taking the test, a feedback survey was given to the test-takers to look into their test-taking experience and feedback on the pilot test.

4.3.3.2. Test construction

i. Task specifications

The needs analysis survey showed that the three main tasks that are frequently used in a business setting are email/letters, memos/notes/messages, and reports/reviews/summaries. Therefore, it was decided that the pilot testing would consist of three tasks: email, memo, and report. Decisions on the specific feature of the task(hereafter "Summary of the BTWS task attributes") were guided by the findings of the needs analysis as well as consideration of the balance of task features (e.g., topic, audience, length, formality, difficulty) across the three tasks.

First, regarding the purpose of writing tasks, the needs analysis showed that expressing one's ideas/opinions, conveying objective information, and demonstrating knowledge/understanding of a topic, are very relevant writing purposes in a business setting. Therefore, the three tasks were designed to fulfill one of the three major writing purposes. Second, since the needs analysis revealed that business writing is almost always performed based on a certain type of input, it was decided that all three tasks involve a certain types of input. The interview confirmed that different types of input, a note/email, table/graph are used for writing tasks. To apply diverse inputs as discovered, the pilot test selected three input types: 1) a short note, 2) a form, and 3) graphs. See the Appendix F for more information about Summary of the BTWS task attributes.

ii. Assessment criteria

Findings of the needs analysis on the important writing skills helped decide assessment criteria for the BTWS. Skills found to be important in the survey included 1) the abilities to organize ideas in a coherent manner, 2) to understand readers' needs, values, and attitudes, 3) to use grammar accurately, and 4) to use appropriate style and tone. Three major assessment criteria selected were Task fulfillment, Idea organization and Language use.

4.3.3.3. Feedback on pilot testing

After pilot testing, participants filled out a short survey for evaluation of the pilot test they had taken. The feedback survey consisted of two major parts: 1) feedback on the test in general and 2) feedback on specific tasks.

i. Overall Evaluation

In general, test-takers evaluated the test positively. As seen in Table 3, 92.3% of the test-takers agreed or strongly agreed that they are satisfied with the test. It was found that many factors contributed to their satisfaction. For example, 100% of the test-takers either agreed or strongly agreed that the three tasks reflected major writing tasks in a workplace setting. The strong agreement confirmed that the tasks contained a high degree of authenticity. In addition, 92.3% of the participants either agreed or strongly agreed that the test seemed to successfully measure writing abilities required in a workplace setting. Besides the authenticity of writing tasks and the relevance of writing skills, test-takers evaluated the assessment criteria, total test-taking time and a computer-delivered format positively. Notably, however, test-takers showed differing opinions on the clarity of test instructions. 57.7% of participants agreed or strongly agreed that the test instructions are clear, whereas 42.3% either disagreed or strongly disagreed with the statement. The finding suggested that test instructions needed to be revised for clarity and explicitness.

Table 3.

Overall evaluation on the pilot test

Items	Strongly Disagree	Disagree	Agree	Strongly Agree	Missing	Total
b1 Generally I am satisfied with the test.	3.8	3.8	69.2	23.1		100
b2 Three tasks (Tasks 1, 2, and 3) reflect major writing tasks required in a workplace setting.	0	0	46.2	53.8		100
b3 The test seems to successfully measure a test-taker's ability to write in a workplace setting.	0	3.8	57.7	34.6	3.8	100
b4 Test instructions are clear	3.8	38.5	34.6	23.1		100
b5 Total test-taking time (60 minutes) is sufficient	0	19.2	38.5	42.3		100
b6 It was interesting to	0	11.5	38.5	50.0		100

perform this test						
b7 A computer-delivered format is appropriate for this writing test	0	3.8	38.5	57.7		100
b8 Assessment criteria used in the test are reasonable	0	3.8	65.4	23.1	7.7	100
b9 My reading ability significantly affects my writing performance on this test	3.8	15.4	42.3	38.5		100

This result was consistent with the findings of the open-ended questions where the often-mentioned weakness of the test was the need to improve instructions. Respondents pointed out more information should be given to guide their responses. Respondent 18 indicates that:

> "The instructions in each task can be improved and clearer . . . The details given sometimes are confusing. Not sure that we have to make it up or only write down the information given".

ii. Feedback on each writing task

1) Task 1-write an email

According to Table 4, 84.7% of the participants either agreed or strongly agreed that they are satisfied with Task 1. According to the table, Task 1 reflected an authentic property of tasks in business settings (100%). Directions for Task 1 were evaluated positively in terms of its clarity (80.8%) and length (96.2%). The results supported that the inputs of Task 1 were clear to understand (73.1%), relevant to the workplace (100%), and had appropriate length (92.3%). In addition, it turned out that test taking time (69.2%) showed relatively low agreement percent, which meant that test taking time might be insufficient to carry out Task 1. However, the required response length (96.2%) was reasonable for participants.

Table 4

Evaluation of Task 1

Items	Strongly Disagree	Disagree	Agree	Strongly Agree	Total
c1 I am satisfied with Task 1	3.8	11.5	38.5	46.2	100
c2 Task 1 is relevant to workplace settings	0	0	34.6	65.4	100
c3 The task directions are clear	3.8	15.4	38.5	42.3	100
c4 The length of task directions is appropriate	0	3.8	57.7	38.5	100
c5 The input (meeting notes) for Task 1 is easy to understand	0	26.9	46.2	26.9	100

c6 The input is relevant to workplace settings	0	0	50.0	50.0	100
c7 The amount/length of input is appropriate	0	7.7	57.7	34.6	100
c8 Test taking time for Task 1 (12 mins) is sufficient	0	30.8	15.4	53.8	100
c9 The required response length (a minimum of 50 words) is reasonable	0	3.8	57.7	38.5	100

Findings from the open-ended questions in the survey reflected respondents'(x8) satisfaction of the task authenticity, which corresponded to the quantitative findings noted above. It was demonstrated by two respondents' comments:

Respondent 1: "We use email a lot!"
Respondent 25: "This type of communication is widely used in the business setting".

Other strengths pointed out included clear instructions (x1) and multiple skills assessed (x1). The latter referred to the ability to comprehend the input prompt correctly and strategically write a reply accordingly. Respondents also offered suggestions in view of test improvement. The most frequent suggestion (x5) made was the need to further elaborate on the context of situation and provided detailed background information. Another interesting point made by one respondent was the input prompt was too long which "affects my writing; too much info in reading". Concerns about the extent to which writing performance was affected by the reading ability were also touched on in the follow-up interviews where interviewees unanimously agreed the two skills go hand in hand.

2) Task 2-write a memo

Table 5 showed participants' evaluation results on Task 2. 84.6% of participants agreed or strongly agreed with the satisfaction of the task 2. The results supported that Task 2 looked like an authentic writing task reflecting the business settings (88.5%). 92.3% of participants evaluated it positively in terms of its clarity of task direction and length of direction (100%). 92.3% agreed or strongly agreed that the inputs of Task 2 seemed clear enough to understand.. Input type for Task 2 was also considered as relevant to workplace (88.5%). 96.2% of participants agreed and strongly agreed that the length of input is appropriate. Moreover, test time (88.5%) and required response length (96.3%) were reasonable to complete Task 2. Table 5.

Evaluation of Task 2

Items	Strongly Disagree	Disagree	Agree	Strongly Agree	Missing	Total
d1 I am satisfied with Task 2	0	7.7	73.1	11.5	7.7	100
d2 Task 2 is relevant to workplace settings	0	11.5	50.0	38.5		100
d3 The task directions are clear	0	7.7	53.8	38.5		100
d4 The length of task directions is appropriate	0	0	57.7	42.3		100
d5 The input (a leave application form) for Task 2 is easy to understand	0	7.7	42.3	50.0		100
d6 The input is relevant to workplace settings	0	11.5	50.0	38.5		100
d7 The amount/length of input is appropriate	0	3.8	50.0	46.2		100
d8 Test taking time for Task 2 (18 mins) is sufficient	0	11.5	38.5	50.0		100
d9 The required response length (a minimum of 120 words) is reasonable	0	7.7	53.8	38.5		100

Findings from the open-ended questions positively indicated respondents'(x4) satisfaction with the task authenticity. It was also pointed out that writing a memo is a good task to measure test-takers' ability to summarize the given input (x2). Nevertheless, concerns about the application or use of memo in current workplace settings were raised. Several respondents (x4) indicated that the memo may have been prevalent in the past but most likely has been replaced by the email because of its extensive and intensive use in the workplace settings.

3) Task 3-write a report

Table 6 presents participants' feedback on Task 3. According to the data analysis, participants also positively evaluated Task 3 since 65.4% of participants were satisfied. 88.5% agreed and strongly agreed that the function of Task 3 reflects the business settings. Clarity and length of directions for Task 3 were evaluated positively, respectively showing that 77% and 96.2% agreed or strongly agreed with the appropriateness of directions for Task 3. The results support that the inputs of Task 3 are relevant to workplace (88.5%), and have appropriate length (88.5%); however, only 61.5% of participants agreed and strongly agreed that the input of Task 3 was clear. Therefore, results suggested the clarity of the input of Task 3 needed further revision.

Furthermore, participants agreed and strongly agreed that test taking time (84.6%) and required response length (96.3%) are reasonable for Task 3.

Table 6.

Evaluation of Task 3

Items	Strongly Disagree	Disagree	Agree	Strongly Agree	Missing	Total
e1 I am satisfied with Task 3	0	26.9	38.5	26.9	7.7	100
e2 Task 3 is relevant to workplace settings	0	11.5	50.0	38.5		100
e3 The task directions are clear	3.8	19.2	30.8	46.2		100
e4 The length of task directions is appropriate	0	7.7	50.0	42.3		100
e5 The input (a chart and a graph) for Task 3 is easy to understand	7.7	30.8	26.9	34.6		100
e6 The input is relevant to workplace settings	0	11.5	57.7	30.8		100
e7 The amount/length of input is appropriate	0	11.5	50.0	38.5		100
e8 Test taking time for Task 3 (30 mins) is sufficient	0	15.4	30.8	53.8		100
e9 The required response length (a minimum of 200 words) is reasonable	0	7.7	57.7	34.6		100

The relatively low agreement on the clarity of inputs of Task 3 reflected respondents' concerns shown in the open-ended responses to improve the clarity of the graphs and instructions. The following two responses indicated the need to enhance the explicitness of input prompts:

Respondent 11: "The graphs are ambiguous".

Respondent 20: "I don't know the content of the first chart. It is a market share, or customer survey share, or something else. I think it should be clear up in the chart".

Additionally, respondents (x2) suggested that the instruction should pinpoint the intended readers as they affect ways that test-takers organize their flow of ideas:

Respondent 5: "I don't know why we have to inform customer about customer's satisfaction. I'm confused that we inform customer who invest with us or not or the customer that use our service +satisfy our service, so it's hard to write down when we're not clear in question".

4.2.2.5. Test revisions after pilot testing

i. Changes in Task 1-write an email

Two major changes were made followed by the results of the pilot testing: a change of input prompt and a requirement of lowest cognitive demand in writing. Meeting notes, used in the pilot test as an input prompt, was found not effective in eliciting the performance of the test-taker's writing ability as test-takers tended to copy the agenda specified in the meeting notes into the response paper. Hence, a shorter input prompt, an email, replaced it to prevent test-takers from copying key phrases and main ideas in the input.

The selection of email as an input was also considered in terms of the extent to which the cognitive domain is required. Intended as the easiest type of writing task, Task 1 aims to involve elementary intellectual activity, namely the skill of knowledge according to Bloom's Taxonomy[2]. This skill requires the simple recognition of facts to illustrate one's ability to list, define, relate, and name, etc. Hence, the input email requires the receiver to write a reply to simply confirm the given information and list items of interests etc.

ii. Changes in Task 2-write a memo

One major change was made to task 2 where the input prompt included instructions of the use of the given form. Findings of the pilot test suggested that examinees tended to copy the instructions in their responses, which makes it difficult to judge examinees' ability to synthesize and analyze the given information. The input prompt was hence revised to avoid giving descriptions of the use of the given form so that examinees are able to demonstrate their comprehension of the input prompt and express it in words.

iii. Changes in Task 3-write a report

The most controversial comment on task 3 was the ambiguity of the graphs and the intended readers. The two graphs were revised so that the first graph, a pie chart, indicated the distribution of product of interests, such as clothing sales amount by regions and mutual fund distribution, while the second graph, a bar chart, shows a comparison of product growth between the current and past years.

Since the report's intended readers are potential or existing clients and shareholders, the information conveyed basically should be positive and promising. Therefore, the revision of Task 3 was made so that the growth rate of product of interest increases in the current year in area where the client's company is located, as compared to the previous or past years.

[2] For more information about Bloom's Taxonomy, see http://www.nwlink.com/~Donclark/hrd/bloom.html

4.4 Claim 4- BTWS is assessed with a valid scoring system

4.4.1. Overview

The purpose of revision of the rating scale, and data analysis to measure raters reliability is to assess a valid scoring system of the BTWS. The rating scale for the BTWS was linked with the CEFR[3] scale, and was evaluated through two rating sessions. Rating scale descriptors were revised to fit the BTWS rating purposes, and definitions of assessment criteria were elaborated by reflecting the raters' perceptions. In addition, raters' performance in the two rating sessions were evaluated in terms of reliability. Based on the results of the two rating sessions, rating scale of the BTWS was finalized and completed.

4.4.2. Linking the BTWS rating scale to the Common European Framework of Reference scale

In view of the target use of the BTWS in a wide range of regions, the BTWS is aimed at alignment with the Common European Framework of Reference (CEFR), an internationally recognized benchmark of language ability, with some adaptations to better reflect the context of the BTWS.

The alignment of the BTWS rating scale to the CEFR was adapted and adjusted through three development stages. The first was the verification of the CEFR assessment criteria that suited the needs of BTWS context by comparing the results of needs analysis and pilot test. The verification results led to a conclusion that the three major linguistic and non-linguistic components that constitutes the CEFR, namely grammar knowledge, pragmatic use and sociolinguistic awareness, corresponded to characteristics essential to second language writing tasks, as found in the findings of needs analysis and pilot test. Some modifications of the CEFR scale were made for the BTWS, in particular, to incorporate three assessment criteria: idea organization, language use, and task fulfillment. Among the three selected criteria, a decision was made according to the pilot rating (see next section) to include assessment of audience awareness in the task fulfillment criterion. The second feature of the BTWS scoring system was to adopt all six levels of proficiency to indicate writing ability, from A1 to C2, as specified in the CEFR. Lastly, 'can-do' statements have been added to describe performance at each level in terms of real skills with language, such as being able to write a report in a coherent manner. The

[3] For more information about CEFR, see http://www.coe.int/T/DG4/Linguistic/CADRE_EN.asp

finalized rating scale is shown in Appendix G of this report.

4.4.3. Rating scale development process

4.4.3.1 First round group rating

A group rating system was followed for the BTWS piloting scoring. A rating model is applied for the pilot rating: four raters, divided into two groups of two raters, participated in the rating. Each group scored half of the essay samples by assigning four analytic scores (Task fulfillment, Language use, Organization, Style awareness) in each task. A reflection log was created for documenting raters evaluation on three writing tasks and assessment criteria. Each rater provided evaluative comments on whether the three proposed tasks are relevant to assess the second writing skills and on whether the descriptors are appropriately phrased to facilitate rating process. In addition, each group provided representative writing samples for each level of proficiency. Results of analysis of the representative writing samples were to help the revision of the BTWS rating scale.

4.4.3.2 Rating scale revision

After finishing the first round of rating, the rating scale was revised on the basis of the four raters' comments in the reflection log. Advantages and disadvantages of three tasks and appropriateness of four assessment criteria were discussed. Based on raters' negotiation and consensus, it was agreed that the number of assessment criteria should be reduced by combining criteria that seems to assess the same features. Hence, three assessment criteria were proposed. Descriptions and definition of new assessment criteria were modified and refined depending on the property of each assessment criteria. A major change, as documented in the current working scale (version 6.2) is as follows:

Berlitz Writing Test_ Analytical Scale Ver.6.2

A. Major changes in Ver.6

1. Reduce the number of assessment criteria from 4 to 3 by incorporating 'Style Awareness', as used in the previous version, into 'Task Fulfillment'
2. Change the naming of 'Organizational Development' to 'Idea Organization'
3. Traits being measured in Task fulfillment: content accuracy, style, understanding of input, communicative goal and audience awareness
4. Traits being measured in Idea Organization: idea development, structure management and use of transition words

B. Descriptions of each assessment criteria:

1) **Task fulfillment** aims to measure examinees' understanding of given input prompts and task requirements and their ability to respond accordingly.
2) **Idea organization** requires examinees to demonstrate their knowledge and use of text structures and supporting details to express intended ideas.
3) **Language use** refers to the ability to recognize and use standard written English, such as vocabulary, sentence and grammar.

4.4.3.3 Second round group rating

The same rater groups who participated in the first round rating were involved in the second round of the pilot rating. In addition, a Berlitz rater also participated in this process as a third rater to facilitate the data analysis. Each rater group rated another half of the essay samples, as opposed to those being rated in the first round of the rating and assigned both analytic scores in each task and a holistic score per essay sample. Criteria used in this round included: Task fulfillment, Idea Organization and Language use. Different from the two rater groups, the Berlitz rater evaluated all 28 essay samples and assigned scores analytically and holistically. Following the same rating procedures as those in the first round of the rating, each rater evaluated the revised assessment criteria and recorded comments on the three tasks in the reflection log (See Appendix D). Representative essay samples at each proficiency level were selected in order to facilitate to finalize the rating scale descriptors and to select prototype samples for each task and each proficiency level.

4.4.3.4 Finalizing the rating scale

Based on the reflection log, the final version of the three assessment criteria and the descriptions were revised to appropriately reflect expected writing performance on each proficiency level (See Appendix G). The prototype writing samples for each proficiency level were selected based on the group consensuses, which were used in the development of training manual. Changes made to the scale are as follows:

A. Task Fulfillment

- Rephrases "Writing Style" to "Style/Audience Awareness"
- Divides the original Task Requirements into two sub categories: Content requirements and Style/audience awareness
- Includes 'reader's effort' as an index to determine A1

B. Idea Organization

- Introduces new trait to be assessed: "plan text that follows conventional structure of essay (i.e. introduction, body and conclusion)".
- Indicates ability to group ideas
- Removes the use of "organizational patterns" as an intended construct being measured

C. Language use

- Introduces new trait to be assessed: " correctness in spelling"
- Modifies the expected syntactic performance in C levels
- Includes 'reader's effort' as an index to determine B1 and levels below

D. Overall

- Attempts to use 'can-do' statement
- Aligns modifiers/quantifiers to those used in the CEF
- Attempts to describe descriptors differently between levels to point out the distinctive features that each level addresses

4.4.4. Pilot data analysis

4.4.4.1. FACETS analysis

Table 7.

Severity and consistency of rater group, three tasks, and assessment criteria

Category	Rater Group			Three Tasks			Assessment Criteria		
	G1	G2	G3	Email	Memo	Report	TF	IR	LU
Severity	-.54	.43	.11	-.05	.04	.00	-.13	-.09	.23
Consistency	1.07	.94	1.01	.94	.94	1.13	1.31	.89	.82
Z score	.8	-.7	.1	-.6	-.7	1.5	3.3	-1.2	-2.2
Agreement	Exact agreement among three rater groups: 37.3%								

Based on the analysis of the second round group rating, the data were analyzed using FACETS and GENOVA. FACETS analysis, developed based on Rasch modeling, provides a consistency (reliability) index and the degree of severity for three facets involved in the pilot test: raters, task and assessment criteria. GENOVA, developed on the grounding of G-Theory, also provides an index of test reliability by measuring the amount of variance among three facets.

Table 7 indicates the results from the FACETS analysis. Five raters in total participated in the second pilot scoring session. Rater 1 and 2 were paired in group 1(G1) and rater 3 and 4 paired in Group 2(G2). The Berlitz rater was classified into Group 3(G3). Each rater scored three writing tasks according to three assessment criteria. The raters were asked to give analytic scores on each assessment criteria and one holistic score per essay.

In Table 7, the second column indicates the inter-rater reliability where the severity level among three groups were almost the same, ranging from -.54 to .43); group 1 was relatively more lenient (-.54) than the other two groups and group 2 was considered relatively harsher in scoring (.43). In terms of intra-rater reliability, three rater groups rated consistently as their infit values for consistency fall within an acceptable range, from .75 to 1.3(McNamara, 1996). This evidence supports that inter- and intra- rater reliability seems to be statistically satisfactory. In addition, results reported that exact agreement among three rater groups was 37.3%, which was considered low.

For three tasks, the levels of difficulty were similar, as shown in column 3. Task 1, write an email, was relatively easy while Task 2, write a memo, was relatively difficult among the three tasks. This result shows a discrepancy between examinees' perception on the difficulty level of each task (See table 8) and raters' perception. Examinees regarded Task 1 as the easiest and Task 3 as the most difficult. Infit values for the three tasks are within acceptable range.

For assessment criteria, in terms of severity level, task fulfillment was the easiest criterion, and language use was regarded as the most difficult assessment criterion. It indicates that raters tended to score leniently on Task Fulfillment; however, raters were likely to score Language Use harshly. In terms of consistency of assessment criteria, two assessment criteria, Idea organization and Language use, exhibit acceptable infit statistics, however, Task Fulfillment presents a sign of inconsistency as misfit statistics reveals (Infit 1.31, Z score 3.3). A possible reason to explain this result is that the description of Task Fulfillment incorporates several aspects of language attributes, which may cause raters to focus on different attributes when rating.

Table 8.

Examinees' perception on relative difficulty level among three tasks

Difficulty Level	Task 1	Task 2	Task 3
1	65.4	23.1	15.4
2	19.2	42.3	30.8
3	15.4	34.6	53.8
Total	100	100	100

4.4.4.2. Analysis of Variance using G-theory

Table 9 presents the results of GENOVA analysis. The third column refers to the percent of variance of each facet (Examinee × Rater × Task × Assessment Criteria). The Examinee facet has variance of 16.9%, and rater facet has 2.7% amount of variance. The outcomes report that tasks and assessment criteria facets do not contain any variance. The results can be interpreted that three facets have very small amount of variance, which indicates that they did not influence factors in scoring. This is a good sign that examinees' ability might be one of strong candidate to affect the scores. Some interaction effects among facets are reported; Interaction between an examinee and tasks facet (13.39%), interaction between a rater and assessment criteria (10.57%), and interaction between an examinee, a rater, and a task (15.79%). Examinees' performance might be affected by types of tasks, and raters' performance might be affected by assessment criteria.

Table 9.

Amount of Variance based on analysis of G-theory

Component	Variance Component		Standard Error
	Variance	Percent	
P (examinee)	0.1605	16.9%	0.0682
R (rater)	0.0256	2.70%	0.0619
I (task)	0.0	0.0%	0.0114
A (assessment criteria)	0.0	0.0%	0.0233
PR	0.0497	5.23%	0.0276
PI	0.1272	13.39%	0.0407
PA	0.0074	0.78%	0.0111
RI	0.0437	4.60%	0.0339
RA	0.1004	10.57%	0.0640
IA	0.0	0.0%	0.0427
PRI	0.1500	15.79%	0.0376
PRA	0.0298	3.14%	0.0154
PIA	0.01883	1.98%	0.0140
RIA	0.020	2.11%	0.0124
PRIA	0.2167	22.81%	0.0207
Total	0.94983	100%	

Table 10 exhibits the G-coefficients on the basis of a D-study and provides a reliability index for each facet. For example, in this pilot test, three rater groups have scored on three test tasks with three assessment criteria, and the G- coefficient is 0.6366, which indicates the overall reliability of the test. This indicates that higher reliability can be achieved when more than three test items are required and more than three raters participate in the operationalized scoring session. The results suggest that more raters, more test items, and more assessment criteria would raise the reliability to 0.8115.

Table 10.

Test reliability index from G-coefficients

Number of each component				
R	I	A	G-coefficient	Phi
1	1	1	0.2111	0.1689
2	2	2	0.4871	0.4214
3	3	3	0.6366	0.5786
4	4	4	0.7216	0.6739
5	5	5	0.7751	0.7358
6	6	6	0.8115	0.7786

Table 11 presents further analysis to extend Table 10 by examining what would happen if the number of elements in each facet were increased (e.g. the addition of more raters). Measures were estimated with the same logic; for instance, when we have more than three assessment criteria (6 assessment criteria), reliability would slightly increase from 0.6366 to 0.6574. Also, if we increase to six raters, the reliability index might be slightly increased from 0.6366 to 0.6984. It is interesting that when we increase the number of test tasks (6 tasks), the reliability would gradually increase from 0.6366 to 0.7379. It is suggested that if we want to obtain higher reliability, it is more reasonable to increase the number of tasks.

Table 11.

Test reliability of three facets based on G-coefficients

# of elements per facet	Assessment Criteria		Rater		Task (item)	
	G	PHI	G	PHI	G	PHI
1	0.5652	0.4819	0.4703	0.3847	0.4110	0.3759
2	0.6171	0.5509	0.5849	0.5139	0.5598	0.5098
3	0.6366	0.5786	0.6366	0.5786	0.6366	0.5786
4	0.6469	0.5935	0.6661	0.6175	0.6836	0.6204
5	0.6532	0.6028	0.6851	0.6434	0.7152	0.6485
6	0.6574	0.6091	0.6984	0.6620	0.7379	0.6689

4.4.4.3 Correlation analysis between BTWS holistic score and TOEFL iBT

Right after the pilot test, 26 participants voluntarily provided their TOEFL scores. Each participant has submitted their TOEFL scores of three different formats (Paper based test, Computer based test, and Internet based test), and these scores were converted to a score standard of iBT TOEFL[4]. The correlation between two tests was 0.442 (n=26) which is significant at .05 level. This indicates a medium association between BTWS and TOEFL iBT scores. Although there is a limitation to generalize results of correlation analysis in that TOEFL iBT scores measures four integrated skills, rather than a discrete language skill, findings suggested to show medium association between the two tests. It is also important to note that the TOEFL is a general academic language proficiency exam, and it was not developed along principles of business LSP, as was the BTWS.

4.4.5. Development of the rater training manual

4.4.5.1. Importance of the training manual

The primary purpose of the training manual is to introduce the overall guidelines of the BTWS test for raters to become familiar with the test. It also could be a useful guidance to facilitate the norming process by providing practice materials and methods in the training manual. A training workshop is a necessary and required process when training raters since it is expected that rater understanding can be enhanced through this workshop.

4.4.5.2 Content of training manual

The training manual consists of two major parts; familiarization, and norming sessions, and some training materials and activities are provided to achieve the goals of training. BTWS training manual introduces useful information for scoring such as the purpose of the BTWS, new assessment criteria, and holistic rating scales for each proficiency level, representative essay samples, and the entire scoring procedures. Furthermore, the training manual included a description of test administration at Berlitz (the client), institutional policy for rating, scoring report system, and a training workshop schedule including the continued workshop program (See Appendix H).

5.0 DISCUSSION

The findings of the four claims suggest that the BTWS is appropriate to the needs of the target test population. Guided by strong theoretical rationales as evidenced in claim 1 and 2, namely a post-Messick validation approach and the framework of Language for Specific Purposes, this study actively used qualitative and quantitative approaches to substantiate findings of the different methods to enhance the robustness of the study. The quantitative evidence as

---

[4] See the website: http://www.ets.org/Media/Tests/TOEFL/pdf/TOEFL_Tips.pdf

collected in the pilot test and rating were confirmed and further explored by the qualitative evidence such as rater dialogue and pilot test-taker voice, both of which contributed considerably to the construct of the second language writing being measured as well as the selection of writing task types and their respective attributes being considered in an attempt to reflect a well-balanced content design. All the empirical evidence collected at different stages of the study, that is the *priori* validity evidence, led support to the claim that the formation of the BTWS reaches high situational and interactional authenticity as was found in the constructs being assessed and selected task types.

The scoring system for the BTWS has been argued in this report as a valid tool for its intended inferences. Adapted from the CEFR scale, the BTWS rating scale has constantly been modified according to the findings of the needs analysis, pilot test and pilot rating. Changes made to the descriptors of each level of proficiency arisen from the 'assessor-oriented' approach (Alderson, 2000) that is intended to guide the rating process and focus on comparing the written text with descriptors on the scale. The results of the pilot ratings where negotiations and consensus between raters were documented served as strong evidence to justify changes on the content of rating scale. Informed by the heuristic paradigm, the qualitative evidence, where raters' voices and reflections were recorded, acknowledges stakeholders' voice during the test development process and helps establish validity argument as raters are most knowledgeable about the context of assessment and able to find coherent interpretations of observation. This approach makes the BTWS contextually specific and sensitive to different test users' interests and their particular value systems.

Nevertheless, more efforts may be placed on establishing rater calibration as the study suggests that variation exists between raters' interpretation of certain assessment criteria. To achieve higher reliability, raters that score the BTWS may be given regular rater training sessions to increase their awareness of the use of rating scale when rating as well as to standardize their interpretation of assessment criteria as is intended in the BTWS rating scale. The G-theory analyses also suggest that adding raters or tasks may benefit the overall reliability of this test.

Lastly, the limitation of this study is that the pilot test-takers recruited in this study are predominately from Asia, including Taiwan, China, India and Korea. Different study findings may be expected if participants come from more diverse ethnic backgrounds.

6.0 RECOMMENDATIONS

Based on the interpretations of the findings, the recommendations are as follows:

1. The degree of clarity of test item prompts had to be enhanced on the basis of the pilot test-takers' feedback.

2. The interaction effect between raters and assessment criteria suggested that clear guidelines for training raters should be provided in order to obtain high rater reliability.

3. According to the results of G-theory, more than two raters should regularly participate in the operational BTWS scoring procedures.

4. Likewise, the G-theory results suggested that if the BTWS can be lengthened, then additional tasks should be added (as opposed to extending the time for the current tasks).

4. An explicit rater training manual needs to be developed as the test rolls out. Based on pilot empirical results, it seems that raters might tend to have a bias on assessment criteria. Raters should be trained to how to understand and interpret the rating scale descriptors and descriptors of three assessment criteria, and this phenomenon should be closely monitored as the test becomes operational.

7.0 CONCLUSION

The study findings suggested that the BTWS can be a reliable and valid test of second language writing ability. The test was designed and elaborated based on empirical evidence. Three test forms and assessment criteria were revised on the ground of comments on participants through the needs analysis and pilot test. The rating scale was completed through several revision processes to reflect stakeholders' perspectives (raters and pilot test-takers). These measures provided adequate guidance for enhancing the test delivery. However, the limitation of this study is that a somewhat small sample size of participants in the needs analysis (n=42) and in the pilot test (n=28); nevertheless, the findings from both measures provided evidence to support the reliability and validity of the current BTWS.

REFERENCES

Alderson, J. C. (1991). Bands and scores. In J. C. Alderson & B. North (Eds.), *Language testing in the 1990s* (pp. 71–86). London: Modern English Publications and the British Council.

Bachman, L.F, (1990). Fundamental Considerations in Language Testing, Oxford University Press

Bachman, L.F., (2005). Statistical Analyses for Language Assessment, Cambridge University Press.

Bachman, L.F., & Palmer, A.D., (1996). Language Testing in Practice, Oxford University Press.

Davidson, F. & Lynch, B. (2002). *Testcraft*. New Haven and London: Yale University London.

Douglas, D.(2000). Assessing language for specific purposes. Cambridge University Press.

Kane, M. (2006). Validation. In Brennan. R. L(Ed.), *Educational measurement* (4th ed., pp. 17-64). New York: American Council on Education and Macmillan.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp 13-103). New York: Macmillan.

McNamara, T. F. (1996). Measuring second language performance. Longman; London and New York.

Miselvy, R., Steinberg, L., & Almond, R. (2003). On the structure of educational assessments. Measurement: Interdisciplinary Research and Perspectives, 1, 3–67.

Toulmin, S. E. (2003). The uses of argument (updated edition). Cambridge, UK: Cambridge University Press.

Weir, C.J. (2005) *Language Testing and Validation: an evidence based approach*, Palgrave MacMillan.

APPENDIX A

Abridged BTWS Test Specification

1. Item type 1: Write an e-mail

General Description (GD): Writing emails is found to be one of the most common written communications in the workplace settings. This portion of the test hence requires test-takers to demonstrate their ability to write a short email to colleagues within the company on a workplace-related matter. To do so, test-takers need to:

- comprehend correctly the given inputs (functional knowledge)
- appropriately display knowledge of the use of grammar, syntax and vocabulary (grammatical knowledge)
- organize the message in a coherent and cohesive fashion (textual knowledge)
- reveal ability to make use of appropriate registers, cultural reference and figure of speech according to audience addressed (style awareness)

Prompt Attributes (PA): A writing prompt in the form of an email is provided (see the sample item below) that explains why the test-taker is writing. In the prompt, three guidelines, along with required response length, are provided that explains what content points test-takers should include. When you write a prompt, also consider the following:

a) *Visual input.*
   - o It refers to an email and is provided on the question paper/screen.
   - o This email should look like a real email that includes subject, date, sender's and receiver's names on the top.
   - o The length of the email should be between 45 to 70 words.

b) *Formality.* The purpose of the email is to exchange short and straightforward information between COLLEAGUES so the language of the prompt is simple, colloquial and less formal.

c) *Cognitive skills.* According to Bloom's taxonomy, the cognitive demand required to fulfill this task involves only the lowest level of Bloom's taxonomy, the skill of "knowledge". Test-takers demonstrate this skill by means of making a list of the main events, writing a list of any pieces of info they can remember, describing what is happening, etc. Therefore, the input prompt (i.e. email) does not ask test-takers to summarize, analyze or evaluate but simply to describe, tell, list, name, request and confirm.

d) *Topic.* Suggested topics are made according to the results of needs analysis and expected cognitive demand required:

Topic	Example
Exchanging info Requesting info	confirm shipment order with clients request status of manual preparation

e) *Guidelines on content points being covered.* Guidelines that explain three required content points are included.
   - First guideline: to confirm the requested info
   - Second guideline: to list the requested items
   - Third guideline: to make a request

Response Attributes (RA): Test-takers will first read the given inputs (i.e. email) and write a reply, in the form of an email, accordingly.

a) *Length of response.* Test-takers need to write a minimum of 50 words. This requirement is shown on the question paper/screen.
b) *Difficulty level.* Given the requirements noted in the PA above, the difficulty level for this task is expected to be the easiest.
c) *Degree of formality.* Given the varying functions of email in daily communication, the expected response is not restricted to any particular business correspondence format. The response can be either formal or informal.

Sample Item (SI):

You have received an email from your colleague regarding the preparation of entertaining the company founder.

Date: Wed 15 Feb 11:20 CST 2009
From: Emma <emma2008@yourcompany.com>
Subject: Company founder's visit
To: Justin <justin1209@yourcompany.com>

Hi, Justin,

I'm writing to ask if there'll be five people, including you, who will give an office tour to Mr. Smith, the company founder, on the day he visits here. Also, could you please let me know where you plan to take Mr. Smith for lunch so I can make a reservation?

Let me know if you have any other special request.

Thank you,
Emma

Write a reply to
- confirm the number of people who will escort Mr. Smith
- list several restaurants you recommend
- request a shuttle bus service from the restaurant to the office

On the next page, please write a minimum of 50 words. You have 12 minutes for this task.

2. **Item type 2-Write a memo**

General Description (GD): Writing a business memo is found to be one of the most frequent writing tasks in the workplace setting. A business memo helps members of an organization tp communicate with each other in an efficient and effective way. Therefore, this portion of the test requires test-takers to demonstrate their ability to write a business memo. To do so, test-takers need to:

- comprehend correctly the given inputs (functional knowledge)
- appropriately display knowledge of the use of grammar, syntax and vocabulary (grammatical knowledge)
- organize message in a coherent and cohesive fashion (textual knowledge)
- reveal ability to make use of appropriate registers, cultural information and figure of speech according to the audience addressed (style awareness)

Prompt Attributes (PA): A business memo is written to perform a specific purpose, such as informing readers of specific information, requesting an action, or reacting to a situation/problem. It is written for a specific audience; therefore it is important to connect its purpose to the interests and needs of the audience. When you write a prompt, consider the following:

a) *Visual input.*
It refers to any kind of form used in workplace, such as an application form, evaluation form, leave application etc.

b) *Targeted audience.*
The targeted audience or readers refer to colleagues or upper management. Clients are not the targeted readers as the memo is intended to be used internally.

c) *Cognitive skills.*
According to Bloom's taxonomy, the cognitive demand required to fulfill this task involves higher levels of Bloom's taxonomy, which refer to skills of "synthesis", and "analysis". Test-takers are expected to demonstrate the skills by putting parts together to form a whole, explaining patterns and meaning and gathering information from given data, etc. Therefore, the prompt requires that test-takers analyze, compare, outline, integrate, and organize etc. The highest cognitive demand, including evaluation, critique and discrimination, is not required in this task.

d) *Guidelines on content points being covered.* Guidelines that explain three required content points are included.

- First guideline: to describe the main purpose(s) of the given input (i.e. a form)
- Second guideline: to describe the form format
- Third guideline: to give general instructions for completing the form

Response Attributes (RA): Test-takers will first read the given inputs (i.e. form), three guidelines for writing a business memo and write a memo accordingly.

a) *Length in response.* Test-takers need to write a minimum of 120 words. This requirement is shown on the question paper/screen.

b) *Difficulty level.* Given the requirements noted in the PA above, the difficulty level for this task is expected to be intermediate, which falls between the other two tasks (i.e. write an email and write a report).

Sample Item (SI):

You are a departmental head. You have been asked to write an announcement memo informing your team members of a new job analysis form. After reviewing the form below, please write the memo to your department.

Your memo should include the following information:

- main purpose(s) of the form
- brief description of its format
- general instructions for completing the form.

Please write a minimum of 120 words. You have 18 minutes for this task.

****

Please complete this form prior to requesting placement of a wanted ad and return it to the department of human resources. This form aids you in the employee selection process.

Job Analysis

Prepared By____________________ **Date** ____________________
Title ____________________ **Department** ____________________

Job Title:	Reports to:

Education/Experience Required

Goals/Objectives of Position

Knowledge/Skills Required

Physical Requirements

Special Problems/Hazards

3. Item type 3-Write a report

General Description: Report of various kinds, such as evaluation report, summary report, financial report etc, is found to be a common type of writing task in the workplace setting. This portion of the test thus requires test-takers to demonstrate their ability to write a report to potential or existing clients on a business-related matter. To do so, test-takers need to:

- comprehend correctly and evaluate analytically the graphic information and express in words(funciatonal knowledge)
- appropriately display knowledge of the use of grammar, syntax and vocabulary (grammatical knowledge)
- organize message in a coherent and cohesive fashion (textual knowledge)
- reveal ability to make use of appropriate registers according to audience addressed (sociolinguistic knowledge)

Prompt Attribute: A writing prompt is provided that explains who the intended reader is, namely potential and existing clients, and what the graphic inputs are. In the prompt, three guidelines, along with the required response length, are provided that explains what content points test-takers should include. When you write a prompt, also consider the following:

a)*Visual/Graphic inputs*.

- o The first graph is a pie chart indicating a total sales/profit/cost amount of product of interests. The other is a bar chart where a comparison of sales/profit/cost growth between current year and the previous year(s) is given.
- o The two graphs will be shown on the question paper/screen.
- o Dos and Don'ts: A report given to potential or existing clients normally conveys a positive message. So the second graph, suggesting the growth trend, should present a positive instead of a negative growth number in client company's area.

b) *Cognitive skills*. According to Bloom's taxonomy, the cognitive demand required to fulfill this task involves higher levels of Bloom's taxonomy, which refer to skills of "synthesis", "analysis" and "evaluation". Test-takers are expected to demonstrate the skills by explaining patterns and meaning, predicting and drawing conclusions and making recommendations, etc. Therefore, the prompt requires test-takers to analyze, infer, integrate, assess, and justify etc.

c) *Topic*. Suggested topics are made according to the results of needs analysis and expected cognitive skills required:

Topic	Example
Work report	progress report, performance report

d) *Guidelines on content points being covered*. Guidelines that explain three required content points are included.

- o First guideline: to describe each given graph
- o Second guideline: to analyze any noticeable trends

- Third guideline: to conclude and recommend according to the graphs

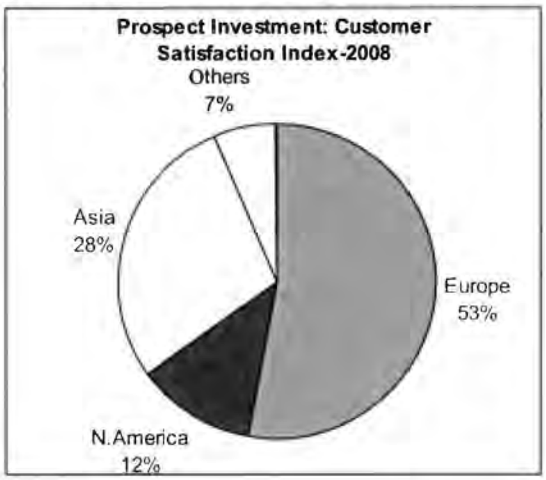
Response Attribute: The test-takers are expected to comprehend the graphic information and express it in words.

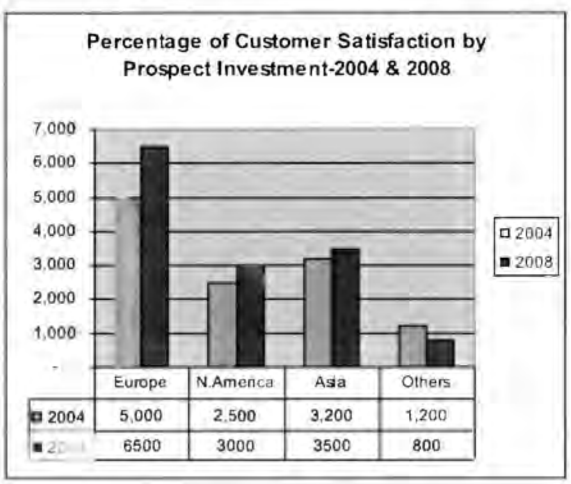
a) Length in response. Test-takers need to write a minimum of 200 words. This requirement is shown on the question paper/screen.

b) Difficulty level. Given the requirements noted above, the difficulty level for this task is expected to be the most difficult.

Sample Item:

You are asked to write a report, to a potential American retailer, analyzing a survey result regarding customer satisfaction in year 2008 and its comparison of year 2004 with year 2008.





This report should include the following information:

- brief descriptions of each chart
- comments on any noticeable trends

- conclusions and/or recommendations based on the data

On the next page, please write a minimum of 200 words. You have 30 minutes for this task.

APPENDIX B

Validity Arguments for the BTWS

	Claims	Assumptions	Evidence	Method of analysis
1	BTWS involves dynamic validation process.	Validation is a dynamic process in which different kinds of evidence are collected and presented. Test specifications are not only a blueprint but also validity narrative.	1. Different types of validity evidence o Needs analysis o Pilot testing o Test-takers' feedback o Consensus-making among raters 2. Test specifications as validity narrative	1. Description of the overall process in which different types of evidence are colleted. 2. The process of due diligence d uring test construction
2	BTWS is developed based on a strong theoretical framework of *Language for Specific Purposes (LSP)*.	A LSP test development requires extensive research on Target Language Use (TLU)	Our major sources of knowledge on TLU, that is, business writing are 1. Needs analysis 2. Competitor analysis	1. Description of needs analysis survey development and administration 2. Document review of existing business writing tests
3	BTWS achieves a high degree of situational authenticity	Needs analysis helps a test developer develop test tasks that successfully reflect TLU.	1. Responses in the needs analysis on the following aspects o writing skills o writing tasks o assessment criteria 2. Test-takers' feedback on the pilot testing	1. Quantitative analysis of the needs analysis survey and test feedback 2. Qualitative analysis of the open-ended questions & interview
4	BTWS is assessed with a valid scoring system	Development of a valid scoring system requires pilot testing and consensus-making process among raters.	1. Our attempts to link the BTWS rating descriptors to CEF 2. Consensus-making process in rating scale development 3. Pilot test score data 4. Development of a rater training manual	1. Explanations of how our descriptors are linked to CEF 2. Explanations of the two-phase rating scale development process, feedback exchange, and constant revisions 3. Quantitative analysis of pilot test scores 4. Explanation of how the rater training manual was developed and what kinds of information are included

APPENDIX C

Quantitative Data Analysis for Needs Analysis

Table 12. Purpose of writing activities

Item	Not relevant at all	Less relevant	Relevant	Very relevant	Total
c1 Write to persuade people	9.5	26.2	42.9	21.4	100
c2 Write to express my own ideas/opinions	0.0	9.5	35.7	54.8	100
c3 Write to demonstrate my knowledge/understanding of a topic	2.4	7.1	31.0	59.5	100
c4 Write to develop an argument	14.3	38.1	35.7	11.9	100
c5 Write to convey objective information	7.1	2.4	23.8	66.7	100
c6 Write to record information for future retrieval	4.8	16.7	33.3	45.2	100

Table 13. Writing inputs

Items	Never	Rarely	Sometimes	Always	Missing	Total
e1.email	0.0	0.0	9.5	90.5	0.0	100
e2.graphic	4.8	23.8	42.9	28.6	0.0	100
e3.paper	2.4	16.7	42.9	38.1	0.0	100
e4.audio	16.7	19.0	31.0	33.3	0.0	100
e5.none	42.9	28.6	16.7	2.4	9.5	100

APPENDIX D

Raters' Comments on the Quality of the Rating Scale (Pilot Rating 2)

1. Suggestions/Comments on Descriptors

Overall

- Reader effort(x1)

I think that in the descriptors for the 3 rated areas we should also include the effect on the reader to the extent possible, e.g. makes it difficult to understand/follow the argument; makes it difficult to identify the main point; causes the reader to go back and re-read the sentence/paragraph; misuse of tenses causes confusion about time, etc..

- Enrich the descriptors(x3)

  - I feel that when looking over all the scores and considering the quality of all the input that the current scale allows us to score more generously than the input really deserves. That is, that in general test takers should be scored approximately1 point less on the scale. In general I think we need to build in additional descriptors that help us to qualify the input more accurately.

  - Descriptors of varying levels may be phrased differently to point out the distinctive features of each level rather than using modifiers/quantifiers only to distinguish levels.

  - Descriptors lack consistency in terms of cognitive demands within and between levels of proficiency. Quantifiers used in the CEF scale: A1: *basic*; limited;
  A2: basic; *simple*; B1: simple; enough, *sufficient*; reasonably accurately
  B2: sufficient; relatively high degree; *appropriately*; clear; clear [p.s. errors are mentioned in this level, mainly that they do not cause misunderstanding];
  C1: good; *flexibly and effectively*; clear; [p.s. errors are mentioned in this level; 'errors are difficult to spot';]; C2: *fully*, consistent.

- Prompt content expected in the response (x1)

For all tasks, how much of the prompt content is expected in the response? For instance, on Task 3: do we expect them to cut-and–paste the graphs into the response? Or for task 1, do we expect them to directly quote the meeting notes? The 'task fulfillment' component descriptors may need editing to address this matter of direct usage of prompt material. For one thing, it is possible to make an essay seem more on task by such borrowings, especially if great attention is given to presentation (e.g. the word wrap around the graphs for Task 3 in #17).

- Core elements of CEF scale (x1)

There are noticeable deviations from the CEF scale, particularly not using the Can-do statement.

1.1. Task Fulfillment

1.1.1) More elaboration on Style Awareness (x2)

- I understand that you incorporated Style into this category? If so, I think it needs to be further developed and include register, politeness. In fact I tend to think that Style should be its own scoring category.
- 'Writing style' may be rephrased to style/audience awareness which indicates not only formality, but also tone used to address different audience.

1.1.2) Evaluation on formality (x2)

- "depending on a writer's interpretation, the email could be either formal or informal. Therefore, regarding writing style, I did not evaluate the formality/informality of writing"
- "It is little hard to look at the formality in Task 1 and task 2. Depending on the understanding of test takers, formality was determined very subjectively"

1.2. Idea Organization

1.2.1) Traits being measured (x2)

- Formal business writing (formal memos, reports and announcements) may have several successful forms but the layout and organization make a big impression. I think this should be better captured in "Idea Organization". To me, this area would include: text organization into paragraphs, main and supporting ideas, introduction/body/conclusion, length of paragraphs, grouping of ideas, all those organizational features that enable the reader to follow the text. These features to my mind are immediately appreciated (or not) when glancing over a passage and predispose the reader either favorably or unfavorably to the text before even one word is read. In addition, the business reader has certain "expectations" about the way a passage should be presented. For example, seeing 10 lines of figures and percentages mixed up in text is unexpected and difficult to wade through.
- It is text structure instead of organizational patterns that should be measured.

1.2.2) Short response length (x3)

- For Task 1, it is hard to judge a score on this criterion "idea organization", because Task 1 required a short response, and there is a difficulty to see the clear idea organization.
- Given the required response length, it is difficult to judge how one develops his idea by using the criteria indicated in the descriptors.
- Required response length is short, so no way to score C2 to any examinee for this criterion.

1.3. Language Use

1.3.1) Features of spelling(x1)

- Should include spelling for all tasks.

1.3.2) Vocabulary & syntactic knowledge (x4)

- It was hard to evaluate *vocabulary* knowledge since many of writers used the same words used in the input. And especially in email, writers often used formulaic expressions.

- It is hard to see the *vocabulary ability and various types of sentences*. So I did look at errors in grammar and misspelling.
- Is using 'a variety of ' sentence structure a prerequisite of a good piece of writing? We're not asking creative writing here.
- ' a variety of' sentence structure? Then few will be scored C2 or C1 for this criterion. This is neither stated in the CEF scale nor informed by the 1st rating results. In other words, function of discrimination is very weak for this criterion.

1.3.3) Traits being measured (x1)

- Descriptors should point out grammar control and use of voc only. Syntax is one aspect of grammatical features.

2. Features that raters attended to [some can be used to inform rater training]

Audience Awareness

- Regarding writing style, since the communicative function is a request, I consider politeness.
- I am wondering how other raters evaluate the use of bullet points in email or memo. I was neutral because it is individual writers' writing style.
- In most essays[for task 3], I could not find relevant audience awareness. This report is to be written for a client, but only few essays (e g., #11) shows audience awareness.
- I want to hear what kinds of features were considered for writing style when other raters review the memo.

Organization

- When writing is written with bullet points, it was hard to evaluate the writers' use of cohesive devices.
- Some essay did miss the writing convention in terms of the writing structure/ idea development. For example, it just shows the clear summary format. In this case, it is little hard to decide the score.

Language use

- See #24 [for task1]– the writer attempted a particularly telegraphic style. I don't think it worked. I am uncertain where to rate that, and so I put it here, in 'use'.

Samples for rater trainings

- We may consider writing a couple of "ideal" responses to each test activity in order to provide a standard. Also, I think that if we could identify and provide sample task responses for each score on the CEF scale, they would be a valuable resource for rater training.

Prompt content

- I am still confused " the degree of the copy of test prompts" when scoring. I think it is o.k with direct copy of some information each task provides, but in the scoring process, I gave higher scores on essay which elaborate basic information. It is reasonable that more modification or paraphrasing is requiring higher level of cognitive process so I think it is fair to give benefit on this kind of essay.

Holistic score

- The instruction in Step 7 to score essays holistically was confusing. There are no holistic descriptors. For the BPI we do have holistic descriptors for speaking as well as the more detailed descriptors for each language area. I think we will also need this for Writing skills.
- Also related to the holistic score, there are no instructions for calculating it. Is it meant to be the mean of the 9 scores in the excel table? Or do you calculate the holistic score for each task, then obtain the mean of the 3 holistic scores? Is a decimal point rounded up or down? In the excel file my holistic score is simply the mean of the previous 9 areas. However, I feel there should an overall score per task, then a single final score.

Rating criteria

- I am still thinking whether it would be good to use different assessment criteria for different tasks like in TOEIC. For example, organization/idea development is least evaluated in the email (Task 1), but it was highly evaluated in the report (Task 2).
- I felt that I am somewhat severe when I grade Tasks 1 and 2 because it was hard to evaluate students' writing proficiency in the checklist-like writing. The easiest part was to evaluate whether they clearly understand the given inputs, but it was hard to evaluate other features. For this reason, I think that Task 3 is a good task that reflects well students' writing proficiency and that we need to provide more room for idea development in other two tasks.

General reflection

- Reflection: It's interesting to see how the majority of test takers do less well on Task 1, which requires a less structured, more interactive style……
- Note: I did not agonize over scoring, therefore not all my scores may stand up in a court of law ☺ but I would also like to go back and compare various scores with the input to ensure consistency up and down the scale. Especially I would revise the scores for task fulfillment for Task 3 in the first tests I scored.

3. Test delivery

- Screen interface

#18 used footnotes in Task 3, but they did not get saved onto the essay (hence, some of that test-taker's content was missing from that essay). If footnoting is OK, we need to implement it with the essay software, and if we permit it with one task, we probably ought to permit it with all tasks.

- Test promotion.

Berlitz may like to address that format does not affect the test score, given that examinees may be new in the job market and have no experience in writing particular tasks, as required in the test.

4. Task Revisions

4.1. Task 1

4.1.1) Copy prompt (x2)

- Test takers often seem to achieve the task goal at least partly by copying phrases from the task instructions. This is something that should be considered when writing test activities and instructions.
- In addition, they ask too much dependency on given input. I agree that these tasks are kind of checklist. We need to develop a task where a writer can develop his/her own ideas based on a given input, not just summarize or review points provided in the input. For example, a give input is email from a client and the task is to respond to the email.

4.1.2) Instructions on formality (x2)

- If we look at the formality and style for task 1, and task 2, we need to give more detail instructions about contextual information between a sender and a receiver.

4.1.3) Distinction btw task 1 & 2 (x2)

- For me, it seems that expected responses for task 1 and task 2 are the same although task 1 and task 2 have a different purpose. We need to think about this point during modification process.
- I feel little confused that based on the test takers' response, their responses look like announcement via formal e-mail message or formal letter, rather than memo. So we need to still consider this point during revision process.

4.2 Task 2

4.2.1) Copy prompt

- Test takers often seem to achieve the task goal at least partly by copying phrases from the task instructions. This is something that should be considered when writing test activities and instructions.

4.3 Task 3

4.3.1) Audience addressed

- The majority of test takers did not realize that this was supposed to be a report addressed to a client and completely missed the point of the activity. I think this task should be modified. It is probably unlikely that you would write a letter to a client analyzing your customer satisfaction and recommending improvements-this is normally dealt with internally. So I think it should either be an internal report or the data should feature the client's satisfaction data and what the company will do to improve it next year.

4.3.2) Clarity of graph

- I think the second graphs looks problematic. Information looks less accurate based on the test takers' interpretation. I think we need to modify "inputs types" or "content of inputs".

APPENDIX E Sample reflection log

	Task 1	Task 2	Task 3
		Comments on rating descriptors	
Task Fulfillment	I understand that you incorporated Style into this category? If so, I think it needs to be further developed and include register, politeness. In fact I tend to think that Style should be its own scoring category.		
Idea Organization	Formal business writing (formal memos, reports and annnouncements) may have several successful forms but the layout and organization make a big impression.I think this should be better captured in "Idea Organization". To me, this area would include: text organization into paragraphs, main and supporting ideas, introduction/body/conclusion, length of paragraphs, grouping of ideas, all those organizational features that enable the reader to follow the text. These features to my mind are immediately appreciated (or not) when glancing over a passage and predispose the reader either favorably or unfavorably to the text before even one word is read. In addition, the business reader has certain "expectations" about the way a passage should be presented. For example, seeing 10 lines of figures and percentages mixed up in text is unexpected and difficult to wade through.		
	Task 1	**Task 2**	**Task 3**
		Comments on task revisions	
	Test takers often seem to achieve the task goal at least partly by copying phrases from the task instructions. This is something that should be considered when writing test activities and instructions.	Test takers often seem to achieve the task goal at least partly by copying phrases from the task instructions. This is something that should be considered when writing test activities and instructions.	The majority of test takers did not realize that this was supposed to be a report addressed to a client and completely missed the point of the activity. I think this task should be modified. It is probably unlikely that you would write a letter to a client analyzing your customer satisfaction and recommending improvements-this is normally dealt with internally. So I think it should either be an internal report or the data should feature the client's satisfaction data and what the company will do to improve it next year.

If you have comments on other aspects, please write them down here.

1. Also related to the holistic score, there are no instructions for calculating it. Is it meant to be the mean of the 9 scores in the excel table? Or do you calculate the holistic score for each task, then obtain the mean of the 3 holistic scores? Is a decimal point rounded up or down? In the excel file my holistic score is simply the mean of the previous 9 areas. However, I feel there should an overall score per task, then a single final score.
2. Reflection: It's interesting to see how the majority of test takers do less well on Task 1, which requires a less structured, more interactive style……

APPENDIX F

Summary of the BTWS task attributes

Task Considerations	**Task 1: Write an Email**	**Task 2: Write a Memo**	**Task 3: Write a Report**
1. Communicative function	Request	Announce	Explain
2. Purpose	Convey objective information	Synthesize given information	Evaluate given information
3. Cognitive demand	Knowledge (lowest)	Analysis and synthesis (intermediate)	Analysis, synthesis and evaluation (highest)
4. Audience	Colleagues	Managers	Clients
5. Prompt input	Email (written)	Form (written)	Chart and graph (graphic)
6. Response length	Minimum 50 words	Minimum 120 words	Minimum 200 words
7. Test taking time	12 minutes	18 minutes	30 minutes
8. Expected formality	Less formal --More formal		
9. Expected level of difficulty	Easy --Difficult		

APPENDIX G

BTWS analytical scale

Level	Task Fulfillment *Essays at this level...*	Idea Organization *Essays at this level...*	Language Use *Essays at this level...*
	Assess examinees' • Understanding of given input • Fulfillment of content requirements • Style/audience awareness	Assess examinees' • Ability to organize text using cohesive devices and to plan text following conventional structure of essay (i.e. introduction, body and conclusion) • Ability to group and present ideas with supporting details/examples	Assess examinees' • Grammatical and syntactic knowledge • Vocabulary use • Correctness in spelling
C2	• Display complete understanding of given input prompts • Fully perform all content requirements • Display competent style/audience awareness	• Plan text fully using conventional structure of essay that includes introduction, body and conclusion and create organized and coherent text consistently using a range of cohesive devices and transition words • Develop a seamless flow of ideas with relevant supporting details and/or examples that show depth and insight.	• Maintain consistent grammatical control and show great flexibility formulating sentence structures • Use vocabulary precisely to give emphasis, to differentiate and to eliminate ambiguity. • Show a consistent degree of competency in spelling
C1	• Display good understanding of given input prompts • Effectively perform almost all content requirements • Display effective style/audience awareness	• Organize text following conventional structure of essay that includes introduction, body and conclusion and effectively show controlled use of cohesive devices and transition words • Develop a smooth flow of ideas with relevant supporting details and/or examples, though a few points may not be fully elaborated.	• Effectively maintain a high degree of grammatical accuracy and have a good command of formulating sentence structures; errors are rare and difficult to spot • Use vocabulary flexibly on professional and general topics • Show correctness in spelling; errors are rare
B2	• Display sufficient understanding of given input prompts • Appropriately perform content requirements even though some points are missing	• Plan text appropriately that shows a clear structure of essay and use a limited number of cohesive devices to link sentences into clear, coherent text, though there may be some 'jumpiness' in a long construction;	• Show a relatively high degree of grammatical control and clearly formulates sentence structures; errors do not cause misunderstanding

	• Display appropriate style/audience awareness even though minor discrepancies are observed	• Appropriately develop ideas with supporting details and/examples, though some points may be less clearly elaborated.	• Have sufficient vocabulary to express viewpoints • Can spell most words correctly
B1	• Display reasonably acceptable understanding of the given input prompt • Attempt to fulfill content requirements, but they are partially performed. • Display reasonably acceptable style/audience awareness, but some inappropriate uses were observed	• Structure text reasonably accurately and can link a series of shorter, discrete simple elements • Display ideas with limited development in depth and thoroughness.	• Use grammar and frequently used sentence patterns reasonably accurately, though readers may be obliged to figure out the meaning after several iterative readings. • Have enough vocabulary to convey meanings on general topics • Can spell correctly the majority of words
A2	• Display basic understanding of the given input prompts • Fulfill content requirements only to a limited extent. • Display basic style/audience awareness; the awareness is either inappropriate or inconsistent	• Have some basic control of structuring text and link groups of words with basic transition words like "and", "but" and "because". • Present ideas that are understandable through reader's efforts but not expanded.	• Use basic grammar and simple sentence structures correctly; mistakes are noticeable without reader's effort • Use basic vocabulary to communicate limited information • Can spell frequently used words correctly only
A1	• Display limited understanding of the given input prompts • Content requirements are rarely fulfilled. • Display limited style/audience awareness; difficult to identify style/audience awareness	• Show limited control of structuring text and link words or groups of words with limited linear connectors such as 'and' or 'then'. • Present ideas that are developed in a limited manner.	• Show only limited control of a few simple grammatical structures and sentence patterns in a memorized repertoire • Have a very limited repertoire of words • Show difficulties with spelling

APPENDIX H
Guidelines for Rater Training

Following are the recommended steps for a rater training workshop for the BTWS:

1. Explain the Structure of the BTWS: What is the BTWS? (See the Specs "introduction" part)

2. Review and practice the three Different Tasks (See the Specs, Part 1, section 1, 2, and 3 for each task type)
   - 3.1 Task 1: Write an e-mail
   - 3.2 Task 2: Write a memo
   - 3.3 Task 3: Write a report

3. Review the assessment Criteria (See the Specs, Part 2, section 1.1)
   - 5.1 Task Fulfillment
   - 5.2 Idea Organization
   - 5.3 Language Use

4. Review the BTWS Rating scale and compare it to the CEFR (See the Specs Part 2 and the Specs Appendix E)
   - 4.1 CEFR Scale
   - 4.2 A1 Level
   - 4.3 A2 Level
   - 4.4 B1 Level
   - 4.5 B2 Level
   - 4.6 C1 Level
   - 4.7 C2 Level

5. Study a prototype Essay for each proficiency level

Prototype essays can be used in the training workshop in order to enhance raters' understanding of making scoring judgment. Based on the consensus between stakeholders (i.e. a trainer and raters), more than two representative essays of varying proficiency levels can be selected within each task and can be used in the lecture and discussion session in the training workshop. Prototype essays can be obtained from the Pilot Testing essay bank, and as the BTWS rolls out, from operational test use.

6. Scoring the Essays – Operational guidelines

This section provides general guidelines on the entire system for methods of scoring examinees' essays. It includes the rating procedures, rating principles, scoring methods, a score report, and rater pairing system.

6.1 Operational Rating Procedure

Examinees' essays are all saved on the Berlitz server and will be assigned to individual raters. Individual raters will access folders on the Berlitz server by using the assigned password. The folder for individual raters contains a rating packet such as a particular test form (set), examinees' essays, rating scale descriptor, and a rating sheet for a score report. The rating results will be automatically submitted to the rating file directly or individual raters will use email to submit the results.

6.2 Rating Principles

6.2.1 Rating Procedure

Before rating, raters should familiarize themselves with the rating scale descriptors for each assessment criteria, each proficiency level, and a particular set of test prompts. The following guidelines describe each scoring step in details.

Step 1: Familiarize yourself with the BTWS rating scale

Three assessment criteria are designed for the BTWS rating scale: Task fulfillment, Idea Organization and Language use. Each assessment criterion has six levels of proficiency, ranging from A1 to C2 (See Appendix E, the Rating Scale).

Step 2: Read writing prompts

Raters should read the test prompts and be familiarized with each task requirements. Pay attention to the context, audience addressed and specific requirements (e.g., word limit and required content for each task).

Step 3: Read examinees' essays

Read Task 1 of all essays, followed by Task 2 and Task 3, respectively. Be sure to familiarize yourself with the specific task requirements before rating.

Step 4: Assign analytic scores

Score Task 1 of all essays, followed by Task 2 and Task 3, respectively. When scoring, assign scores according to the three assessment criteria. Mark the scores on the official rating sheet provided. If you have difficulty in determining scores, look at the rating scale again and make the best judgment you can. Make a note of reasons that result in difficulty in determining scores. The follow-up rater training session will allow you to discuss issues that you encounter when assigning scores.

6.2.2 The Scoring Method and System

Raters should assign three analytic scores across three tasks, and the holistic score is calculated by averaging three analytic scores. The result will be used in a score report with descriptions of each proficiency level for examinees.

6.2.3 Rater Pairings
Berlitz informs and assigns a model for rater pairings via individual emails. Two raters participate in the first rating and the third rater's participation is necessary when a discrepancy of one point occurs between scores assigned by the first two raters.

6.3 Institutional rating policy (The Dos and DON'Ts of rating)

- Before starting rating, familiarize yourselves with the overall scoring procedures.
- Fully understand the test prompts and directions before scoring.
- Make scoring judgment according to the rating scale descriptors.
- During rating, the degree of formality of examinees' responses is not the critical rating point.
- Try to find out the advantages and disadvantages of each essay. It is helpful to determine the broad divisions first (A, B, and C) and then the six proficiency levels.
- Do not compare essays during rating, and use criteria that the rating scale provides.
- Do not use your internal standards(e.g. your belief in what constitutes a good writing) that are not included in the rating scale If you have problems in scoring, review the training manual and prototype essays for each level.

7. Other issues
This section provides general information about the rater workshop program such as the role of a trainer, workshop schedule, and continued education program.

7.1 Role of a Trainer
The primary role of a trainer, as a facilitator, is to train new and experienced raters by providing practice sessions during the workshop. Also, a trainer encourages raters to conduct their task and gives advice to solve their problems in scoring. Additionally, a trainer plays a role to deliver raters' messages/concerns to the testing agency.

7.2 Training Workshop Administration

7.2.1 General information for workshop
Berlitz provides a regular training workshop in order to train newly recruited raters. The primary audience is new raters, and the training workshop lasts for four days. After participating in all the workshop activities, candidates will become official raters.

7.2.2 Continued Education

Berlitz provides a continued education program for experienced raters in the form of on-site and on-line rating program. The BTWS trainers or official BTWS raters can request the continued education program. There are two purposes of the continued education: refreshment before rating, or extensive discussions for the problem-solving. The continued education is a useful program for experienced raters to review the rating principles and retrain their skills before the rating. Furthermore, if raters have problems in scoring, then it can be a place to discuss their concerns and share the solutions with each other.

CITY OF AUSTIN
PURCHASING OFFICE
SUPPLEMENTAL PURCHASE PROVISIONS

The following Supplemental Purchasing Provisions apply to this solicitation:

1. **EXPLANATIONS OR CLARIFICATIONS** (reference paragraph 5 in Section 0200)

   All requests for explanations or clarifications must be submitted in writing to the Purchasing Office by 2:00p.m. on 11/4/2012 via email at Jeffery.dilbert@austintexas.gov or FAX at 512-974-2388

2. **INSURANCE.** Insurance is required for this solicitation.

   A. General Requirements. See Section 0300, Standard Purchase Terms and Conditions, paragraph 32, entitled Insurance, for general insurance requirements.

      i. The Contractor shall provide a Certificate of Insurance as verification of coverages required below to the City at the below address prior to contract execution and within 14 calendar days after written request from the City. Failure to provide the required Certificate of Insurance may subject the Offer to disqualification from consideration for award

      ii. The Contractor shall not commence work until the required insurance is obtained and until such insurance has been reviewed by the City. Approval of insurance by the City shall not relieve or decrease the liability of the Contractor hereunder and shall not be construed to be a limitation of liability on the part of the Contractor.

      iii. The Contractor must also forward a Certificate of Insurance to the City whenever a previously identified policy period has expired, or an extension option or holdover period is exercised, as verification of continuing coverage.

      iv. The Certificate of Insurance, and updates, shall be mailed to the following address:

      City of Austin Purchasing Office
      P. O. Box 1088
      Austin, Texas 78767

   B. Specific Coverage Requirements. The Contractor shall at a minimum carry insurance in the types and amounts indicated below for the duration of the Contract, including extension options and hold over periods, and during any warranty period. These insurance coverages are required minimums and are not intended to limit the responsibility or liability of the Contractor.

      i. Worker's Compensation and Employers' Liability Insurance. Coverage shall be consistent with statutory benefits outlined in the Texas Worker's Compensation Act (Section 401). The minimum policy limits for Employer's Liability are $100,000 bodily injury each accident, $500,000 bodily injury by disease policy limit and $100,000 bodily injury by disease each employee.

         (1) The Contractor's policy shall apply to the State of Texas and include these endorsements in favor of the City of Austin:
            (a) Waiver of Subrogation, Form WC 420304, or equivalent coverage
            (b) Thirty (30) days Notice of Cancellation, Form WC 420601, or equivalent coverage

      ii. Commercial General Liability Insurance. The minimum bodily injury and property damage per occurrence are $500,000 for coverages A (Bodily Injury and Property Damage) and B (Personal and Advertising Injury).

         (1) The policy shall contain the following provisions:

(a) Contractual liability coverage for liability assumed under the Contract and all other Contracts related to the project.
(b) Contractor/Subcontracted Work.
(c) Products/Completed Operations Liability for the duration of the warranty period.
(d) If the project involves digging or drilling provisions must be included that provide Explosion, Collapse, and/or Underground Coverage (X,C,U).

(2) The policy shall also include these endorsements in favor of the City of Austin:
(a) Waiver of Subrogation, Endorsement CG 2404, or equivalent coverage
(b) Thirty (30) days Notice of Cancellation, Endorsement CG 0205, or equivalent coverage
(c) The City of Austin listed as an additional insured, Endorsement CG 2010, or equivalent coverage

iii. Business Automobile Liability Insurance. The Contractor shall provide coverage for all owned, non-owned and hired vehicles with a minimum combined single limit of $500,000 per occurrence for bodily injury and property damage. Alternate acceptable limits are $250,000 bodily injury per person, $500,000 bodily injury per occurrence and at least $100,000 property damage liability per accident.

(1) The policy shall include these endorsements in favor of the City of Austin:
(a) Waiver of Subrogation, Endorsement TE 2046A, or equivalent coverage
(b) Thirty (30) days Notice of Cancellation, Endorsement TE 0202A, or equivalent coverage
(c) The City of Austin listed as an additional insured, Endorsement TE 9901B, or equivalent coverage.

C. Endorsements. The specific insurance coverage endorsements specified above, or their equivalents must be provided. In the event that endorsements, which are the equivalent of the required coverage, are proposed to be substituted for the required coverage, copies of the equivalent endorsements must be provided for the City's review and approval.

3. TERM OF CONTRACT

A. The Contract shall be in effect for an initial term of 36 months and may be extended thereafter for up to one additional 36 month period, subject to the approval of the Contractor and the City Purchasing Officer or his designee.

B. Upon expiration of the initial term or period of extension, the Contractor agrees to hold over under the terms and conditions of this agreement for such a period of time as is reasonably necessary to re-solicit and/or complete the project (not to exceed 120 days unless mutually agreed on in writing).

C. Upon written notice to the Contractor from the City's Purchasing Officer or his designee and acceptance of the Contractor, the term of this contract shall be extended on the same terms and conditions for an additional period as indicated in paragraph A above. A price increase, subject to the provisions of this Contract, may be requested by the Contractor (for each period of extension) for approval by the City's Purchasing Officer or his designee.

THIS IS A 36 MONTH CONTRACT.

FIRM PRICES ARE TO BE SUBMITTED FOR THE FIRST THIRTY-SIX (36) MONTH PERIOD

4. QUANTITIES

The quantities listed herein are estimates for the period of the Contract. The City reserves the right to purchase more or less of these quantities as may be required during the Contract term. Quantities will be as needed and specified by the City for each order. Unless specified in the solicitation, there are no minimum order quantities.

5. **INVOICES and PAYMENT** (reference paragraphs 12 and 13 in Section 0300)

A. Invoices shall contain a non-duplicated invoice number and the information required in Section 0300, paragraph 12, entitled “Invoices.” Invoices received without all required information cannot be processed and will be returned to the vendor.

Invoices shall be mailed to the below address:

	City of Austin
Department	Human Resources Department
Attn:	Accounts Payable
Address	PO Box 1088
City, State Zip Code	Austin, TX 78767

B. The Contractor agrees to accept payment by either credit card, check or Electronic Funds Transfer (EFT) for all goods and/or services provided under the Contract. The Contractor shall factor the cost of processing credit card payments into the Offer. There shall be no additional charges, surcharges, or penalties to the City for payments made by credit card.

6. **LIVING WAGES AND BENEFITS (applicable to procurements involving the use of labor)**

A. In order to help assure low employee turnover, quality services, and to reduce costs for health care provided to uninsured citizens, the Austin City Council is committed to ensuring fair compensation for City employees and those persons employed elsewhere in Austin. This commitment has been supported by actions to establish a “living wage” and affordable health care protection. Currently, the minimum wage for City employees is $11.00 per hour. This minimum wage is required for any Contractor employee directly assigned to this City Contract, unless Published Wage Rates are included in this solicitation. In addition, the City may stipulate higher wage rates in certain solicitations in order to assure quality and continuity of service.

B. Additionally, the City provides health insurance for its employees, and for a nominal rate, employees may obtain coverage for their family members. Contractors must offer health insurance with optional family coverage for all Contractor employees directly assigned to this contract. Proof of the health care plan shall be provided prior to award of a Contract. In addition, an insurance certificate for Workers’ Compensation Insurance Coverage must be provided if required by the solicitation.

C. The City requires Contractors submitting Offers on this Contract to provide a signed certification **(see the Living Wages and Benefits Contractor Certification included in the Solicitation**) with their Offer certifying that all employees directly assigned to this City Contract will be paid a minimum living wage equal to or greater than $11.00 per hour and are offered a health care plan. The certification shall include a list of all employees directly assigned to providing services under the resultant contract including their name and job title. The list shall be updated and provided to the City as necessary throughout the term of the Contract.

D. The Contractor shall maintain throughout the term of the resultant contract basic employment and wage information for each employee as required by the Fair Labor Standards Act (FLSA). Basic employment records shall at a minimum include:

i. employee's full name, as used for social security purposes, and on the same record, the employee's identifying symbol or number if such is used in place of name on any time, work, or payroll records;
ii. time and date of week when employee's workweek begins;
iii. hours worked each day and total hours worked each workweek;
iv. basis on which employee's wages are paid;
v. regular hourly pay rate;
vi. total daily or weekly straight-time earnings;
vii. total overtime earnings for the workweek;
viii. all additions to or deductions from the employee's wages;
ix. total wages paid each pay period; and
x. date of payment and the pay period covered by the payment.

E. The Contractor shall provide with the first invoice and as requested by the Department's Contract Manager, individual Employee Certifications **(see the Living Wages and Benefits Employee Certification included in the Solicitation)** for all employees directly assigned to the contract containing:

i. the employee's name and job title;
ii. a statement certifying that the employee is paid at a rate equal to or greater than the Living Wage of $11.00 per hour;
iii. a statement certifying that the employee is offered a health care plan with optional family coverage.

Employee Certifications shall be signed by each employee directly assigned to the contract.

A. Contractor shall submit employee certifications quarterly with the respective invoice to verify that employees are paid the Living Wage throughout the term of the contract.

G. The Department's Contract Manager will periodically review the employee data submitted by the Contractor to verify compliance with this Living Wage provision. The City retains the right to review employee records identified in paragraph D above to verify compliance with this provision.

7. NON-COLLUSION, NON-CONFLICT OF INTEREST, AND ANTI-LOBBYING

A. On November 10, 2011, the Austin City Council adopted Ordinance No. 20111110-052 amending Chapter 2.7, Article 6 of the City Code relating to Anti-Lobbying and Procurement. The policy defined in this Code applies to Solicitations for goods and/or services requiring City Council approval under City Charter Article VII, Section 15 (Purchase Procedures). During the No-Contact Period, Offerors or potential Offerors are prohibited from making a representation to anyone other than the Authorized Contact Person in the Solicitation as the contact for questions and comments regarding the Solicitation.

B. If during the No-Contact Period an Offeror makes a representation to anyone other than the Authorized Contact Person for the Solicitation, the Offeror's Offer is disqualified from further consideration except as permitted in the Ordinance.

C. If a Respondent has been disqualified under this article more than two times in a sixty (60) month period, the Purchasing Officer shall debar the Offeror from doing business with the City for a period not to exceed three (3) years, provided the Respondent is given written notice and a hearing in advance of the debarment.

D. The City requires Offerors submitting Offers on this Solicitation to provide a signed Section 0810, Non-Collusion, Non-Conflict of Interest, and Anti-Lobbying Affidavit, certifying that the Offeror has not in any way directly or indirectly made representations to anyone other than the Authorized Contact Person during the No-Contact Period as defined in the Ordinance The text of the City

Ordinance is posted on the Internet at: http://www.ci.austin.tx.us/edims/document.cfm?id=161145

8. **NON-SOLICITATION**

A. During the term of the Contract, and for a period of six (6) months following termination of the Contract, the Contractor, its affiliate, or its agent shall not hire, employ, or solicit for employment or consulting services, a City employee employed in a technical job classification in a City department that engages or uses the services of a Contractor employee.

B. In the event that a breach of Paragraph A occurs the Contractor shall pay liquidated damages to the City in an amount equal to the greater of: (i) one (1) year of the employee's annual compensation; or (ii) 100 percent of the employee's annual compensation while employed by the City. The Contractor shall reimburse the City for any fees and expenses incurred in the enforcement of this provision.

C. During the term of the Contract, and for a period of six (6) months following termination of the Contract, a department that engages the services of the Contractor or uses the services of a Contractor employee will not hire a Contractor employee while the employee is performing work under a Contract with the City unless the City first obtains the Contractor's approval.

D. In the event that a breach of Paragraph C occurs, the City shall pay liquidated damages to the Contractor in an amount equal to the greater of: (i) one (1) year of the employee's annual compensation or (ii) 100 percent of the employee's annual compensation while employed by the Contractor.

9. **ECONOMIC PRICE ADJUSTMENT**

A. Prices shown in this contract shall remain firm for the first 36 month period of the contract. After that, in recognition of the potential for fluctuation of the Contractor's cost, a price adjustment (increase or decrease) may be requested by either the City or the Contractor subject to the following considerations:

B. Price Increases

i. Requests for price increases must be made in writing and submitted to the appropriate Buyer in the City's Purchasing Office. The letter must be signed by a person with the authority to bind the Contractor contractually, shall reference the contract number, and include the following documentation:

(1) an itemized, revised price list with the effective date of the proposed increase;

(2) copies of the documentation provided by the manufacturer regarding the proposed price increase if the contractor is not the manufacturer of the products. If the Contractor is the manufacturer of the products, a letter so stating must be provided;

Contractor shall submit, as a part of the request for increase, the version of the Producer Price Index (s) (Table 5 - Producer price indices for the net output of selected industries and their products or services) Industry/Product name/code (the "Index") current as of the date of the Contractor's Offer; and a copy of the index for the most current period.

(3) Proposed price increases must be solely for the purpose of accommodating increases in the Contractor's costs for the products or services provided. Prices for products or services unaffected by verifiable cost trends shall not be subject to change.

ii. Requests for price increases must be made in writing and submitted to the appropriate Contract Manager prior to each yearly anniversary date of contract. Prices will only be considered for an increase at that time. Once received, the City will have 30 calendar days to review and approve/disapprove the requested increase. Should the City not agree with the requested increase, Contractor may either maintain the prices currently in effect, negotiate an acceptable increase with the City or terminate the contract.

iii. Since the perceived need for price increases may be due in whole or in part to factors other than index changes, the City may consider approving fully-documented increase requests which, in the Contractor's opinion, justify price increases for one or more line items in the contract. If index changes are responsible in part for the requested change, those changes shall be documented as previously described above.

C. Price Decreases

i. Proposed price decreases may be offered to the City at any time, and become effective upon acceptance by the City unless a different effective date is specified by the Contractor. Request for price decreases by the City will be based on the same documentation as price increase request. Price decrease offers may also be subject to negotiation.

ii. Price decreases based on relevant factors may be requested by the City at any time. Such requests shall be accompanied by a complete statement of the City's justification for the request. The Contractor shall have 30 calendar days to respond to the City's request. Following receipt of the Contractor's agreement with the requested decrease, the City may implement the decrease at any time. Should the Contractor not agree with the requested decrease, the City may either maintain the prices currently in effect, negotiate with the contractor, or terminate the contract.

10. **INTERLOCAL PURCHASING AGREEMENTS** (applicable to competitively procured goods/services contracts).

A. The City has entered into Interlocal Purchasing Agreements with other governmental entities, pursuant to the Interlocal Cooperation Act, Chapter 791 of the Texas Government Code. The Contractor agrees to offer the same prices and terms and conditions to other eligible governmental agencies that have an interlocal agreement with the City.

B. The City does not accept any responsibility or liability for the purchases by other governmental agencies through an interlocal cooperative agreement.

11. **CONTRACT MANAGER**

The following person is designated as Contract Manager, and will act as the contact point between the City and the Contractor during the term of the Contract:

Holly Moyer

512-974-3216

*Note: The above listed Contract Manager is not the authorized Contact Person for purposes of the **NON-COLLUSION, NON-CONFLICT OF INTEREST, AND ANTI-LOBBYING Provision** of this Section; and therefore, contact with the Contract Manager is prohibited during the no contact period.