BRIGHT STANDARD ENGLISH TEST

VALIDITY & RELIABILITY REPORT

Bright Standard English Test
Validity & Reliability Report

Table of Contents

Section 1: Introduction

About this report

This report illustrates the content and technical details of Bright Standard English Test, BSET, an English language proficiency exam offered by Bright Online. Bright Online developed the exam to provide both test-takers and organizations with an accessible and accurate test of English proficiency.   The primary audience is test users or stakeholders who would like to make informed decisions regarding test usage. It is intended to provide decision-makers with the necessary information to confidently use Bright Standard English Test.

The report shares detailed information on:

- the theoretical framework for language assessment
- the content of BSET
- how the test is scored and reported
- the development of the exam
- the measurements of validity and reliability

This report is divided into seven sections. Section 1 is an introduction that gives an overview of Bright Online and BSET. Section 2 provides the theoretical framework of language assessment and how it relates to BSET, while Section 3 gives a comprehensive description of BSET. Section 4 describes the exam procedure and Section 5 explains the scoring and proficiency levels. Section 6 discusses the test development process and Section 7 lays out the reliability, validity, authenticity and washback of BSET.

About Bright Online

Bright Online, led by professional global educators, is an online provider of English language curriculum, instruction and assessments. Founded, as Bright Learning, in 2017 to solve the

disconnect between student needs and language school offerings by providing online English courses. Expanded in 2020 as Bright Online LLC to include all brands and programs under one entity; Bright Learning for adult English courses and certifications, Bright Kids for English courses specifically for children and Bright Exams for English proficiency assessments. Bright Online's mission is to brighten learners' futures by helping them gain the English language skills and credentials they need to succeed.

Bright Online is an international affiliate of AERA (American International Research Association) and an institutional member of ILTA (International Language Testing Association) and ICC (The International Language Association). Additionally, Bright Standard English test holds a Quality Assurance certification from ICC. Our researchers are also members of EALTA (European Association for Language Testing and Assessment), ELT Society (English Language Testing Society), and ALTE (Association of Language Testers in Europe). Our courses and proficiency exams are accredited by IAQLS (International Accreditation for Quality Language Services), EAHEA (European Agency for Higher Education) and ECICEL (European Centre of Independent Certification in E-Learning).

About Bright Standard English Test

Bright Standard English Test was developed by a team of English language teaching, learning and assessment professionals to provide English language learners with an accessible English exam that reflects the use of English in academia and professional settings. The content of BSET is designed to represent the academic and professional experiences of students and employees. Furthermore, it was also developed to accurately reflect candidates' English language proficiency thus providing institutions, organizations and companies with true and reliable depictions of candidates' skills and abilities. Candidates' English language proficiency results from BSET are valid for both educational and employment purposes. The developers understood the importance of evaluating both candidates' receptive (reading and listening) and productive (writing and speaking) skills and made a distinction between these skills within the test. The developers also recognized that English language learners need to be able to use

English in real-world communicative situations. In this regard, the exam mirrors the types of tasks that candidates encounter in communicative situations in English in all four skill areas: speaking, listening, reading, and writing. BSET reflects best practices in current language teaching and learning, and best practices in methodology with a focus on the Communicative Approach. BSET acknowledges that English is a Lingua Franca and its language use tasks revolve around daily, academic and professional life with non-native and native English speakers.

Section 2: Theoretical Framework for Language Assessment

Language skills are generally grouped into four separate categories by language educators and researchers: listening, speaking, reading, and writing. This distinct categorization of language skills into four categories is easily understood as each element relates to a set of activities. Furthermore, language researchers and educators commonly make a distinction between language skills and language knowledge. Language knowledge often refers to vocabulary and grammar. This knowledge is necessary and is utilized across all four language skills. Vocabulary and grammar knowledge is necessary to be proficient in reading, writing, speaking and listening. The categorization of language skills and language knowledge has formed the basis of research in language assessment for the past five decades (Chapelle, 2008) and forms the foundation of the design and structure of Bright Standard Test of English. The four sections of BSET (reading, writing, listening and speaking) align with the traditional categorization of language skills while language knowledge is assessed throughout all four sections. Research shows that when all skills are assessed, the assessment is both fairer (to the test-taker) and more accurate (for the test-user) (DE Powers, 2010).

While language educators, researchers and test developers have traditionally delineated language skills into four categories, it is important to note that language skills are not isolated from each other. Modern language theories, such as the integrative theory, have placed emphasis on the interconnectedness of language skills and knowledge within a meaningful context. In authentic communicative interactions, skills and knowledge are not separate entities, but are rather used

simultaneously. BSET follows traditional categorization while fully embracing the integrative theory that each of the skills is integrated with other skills in practice and with the knowledge of grammar and vocabulary. For example, listening assessments require examinees to also read listening items, make meaning from grammar, know vocabulary words and understand sentence structure.

In addition, language skills and language knowledge are all embedded in a communicative context. This communicative context and situation must be considered when evaluating English language proficiency levels (Association of Language Testers in Europe, 2011; Bachman, 1990). Furthermore, Bachman and Palmer (2010) state that language assessments should be designed to engage candidates with tasks that closely mirror real-life situations that candidates will find themselves in outside of the assessment. Assessment developers should define target language use domains, "a specific setting outside the test itself that requires the test taker to perform language use tasks" (Bachman, 2010, p.60). In this regard, BSET embeds language knowledge (grammar, vocabulary) and language skills (reading, writing, speaking, listening) in the context and setting of academia and professional workplaces. BSET emphasizes the authentic use of language in academic and business settings and is designed to represent meaningful real-world language use. It was designed in accordance with best practices in language assessment based on well-known theoretical frameworks. BSET effectively evaluates candidates' language proficiency across language skills and knowledge in the context of common situations in academia and professional settings.

Section 3: Description of Bright Standard English Test

Overview of Typical Test Taker

BSET is designed for candidates aged 16 and above to assess their English language abilities in all four skills and overall understanding of English. It is designed to provide assessment options for English language learners from levels A1 to C2 who are involved in education, training or employment. The description of test takers is comprehensive as BSET is meant to be accessible to a broad range of potential test-users across a range of contexts.

Test Purpose

BSET tests and evaluates English proficiency levels for learners of English. No particular culture, first language background or specialized knowledge of candidates was targeted during test development. The test content and design is appropriate for candidates from a variety of cultural and linguistic backgrounds. BSET is used for a wide-range of purposes, including both in educational and professional contexts. The test is provided both directly to a broad-spectrum of candidates and to organizations globally.

BSET is commonly used for:

- Identifying candidates for admission into educational programs
- Making level-appropriate placements of students in language learning and training programs
- Identifying strengths and weaknesses of students to inform needs-based curriculum and course instruction
- Evaluating and tracking pre- and post-course progress
- Verifying English proficiency for employment purposes

Candidates may use the results of BSET for academic and professional requests for evidence of their English language abilities. Bright Standard English Tests' accurate and reliable evaluations of candidates' English language proficiency across all skills allows educators, administrators and employers to make confident, evidence-based decisions regarding enrollment, placement, employment and promotion.

Target Language Use (TLU) Domain

During the test design process, target language use (TLU) domains of potential candidates were identified and subsequently laid the foundation for test specifications and item writing. The exam evaluates candidates' ability to partake in authentic communication in the personal, public,

educational, and occupational domains. Adhering to the target language use domains throughout test design guarantees that BSET results remain meaningful to both candidates and stakeholders. The CEFR describes the target language use domains in the following ways:

Personal	Public	Educational	Occupational
"the person concerned lives as a private individual, centred on home life with family and friends, and engages in individual practices such as reading for pleasure, keeping a personal diary, pursuing a special interest or hobby, etc."	"the person concerned acts as a member of the general public, or of some organisation, and is engaged in transactions of various kinds for a variety of purposes"	"the person concerned is engaged in organised learning, especially (but not necessarily) within an educational institution"	"the person concerned is engaged in his or her job or profession"

The situations that occur within domains can further be described in terms of locations, institutions, persons, objects, events, operations and texts. The TLU tasks involve candidates using English to communicate in real-world situations at school, work and in daily life.   TLU contexts are those in which secondary students, university students and adults using English will find themselves.

BSET Personal Domain Situations (non-exhaustive)						
Locations	Institutions	Persons	Objects	Events	Operations	Texts
house, apartment, garden, the yard, the shed, the garage, the seaside, the countryside, villages, the gym, the park, the beach, the forest, mountains, the desert, towns, cities, suburbs,	family, community, extended family, social networks, recreation centers, health centers, community centers, government organizations, non-profits	child, parents, offspring, baby, grandparents, grandchildren, cousins, uncles, aunts, in-laws, girlfriend, boyfriend, friends, relatives, nephews, nieces,	decorations and furniture, household goods and appliances, toiletries, sport equipment, clothing, pets and their items, tools, arts and crafts supplies, flora and fauna	family events, parties and celebrations, holidays, vacations, incidents and accidents, emergencies and natural phenomena, sporting events	daily living routines, taking care of a household, leisure activities, hobbies, entertaining, sports and games	bills and mail, leisure reading material, academic reading material, emails, personal letters, social media posts, news, warranties, instruction manuals,

commuter towns, outskirts		siblings,ancestors, neighbors, acquaintances				recipes, receipts, menus

BSET Public Domain Situations (non-exhaustive)						
Locations	**Institutions**	**Persons**	**Objects**	**Events**	**Operations**	**Texts**
public spaces, supermarkets, malls, shops, farmers' markets, libraries, clinics, hospitals, dental clinics, sports stadiums, gyms, restaurants, cafes, coffee shops, bakeries, community centers, museums, aquariums and zoos, amusement parks, gas stations, airports, train stations, theaters	government organizations and departments, non-profits, public health organizations, cultural centers, embassies and consulates	members of the government, members of the military, artists, actors and actresses, sports players and fans, members of the public, waiters and waitresses, public officials, clerks, administrators, hotel managers, receptionists, sales people, civil servants	money, wallet, purse, forms, goods, food and drinks, identification, documents, applications, invoices, receipts, websites, certificates,	incidents and accidents, sporting events and competitions, public meetings, concerts and entertainment, press conferences, trials, ceremonies, conferences and exhibitions,	purchasing and using public services, using medical services, using public and private transportation, attending events, fundraising, volunteering	notifications, announcements, labels and packaging, brochures, advertisements, programs, menus, books, fines, information packs

BSET Educational Domain Situations (non-exhaustive)						
Locations	**Institutions**	**Persons**	**Objects**	**Events**	**Operations**	**Texts**
schools, colleges, universities,	schools, colleges, universities,	teachers, professors, lecturers, staff,	reading and writing material,	conference, forums, competitions,	attending lessons and lectures, essay	academic papers, articles and journals,

lecture halls, classrooms, laboratories, workshops, training centers, libraries	academic organizations, educational bodies	counselors, mentors, facilitators, assistants, parents, students, classmates	computers, laptops, phones, school bags and supplies	trips and excursions, special events, concerts, graduation and ceremonies, athletics	writing, doing homework, taking exams, attending seminar, debating, discussing, note-taking, experimenting, tutoring, participating in events and ceremonies	reference material, textbooks, timetables, course outlines, dissertations, bios, websites

BSET Occupational Domain Situations (non-exhaustive)						
Locations	**Institutions**	**Persons**	**Objects**	**Events**	**Operations**	**Texts**
places of employment, offices, firms, clinics, studios, workshops, hospitals, hotels, resorts, centers, museums, stores, businesses, government buildings	professional organizations and bodies, government organizations and departments, affiliations	employees, employers, coworkers, clients, guests, customers, partners, boards, mentors, advisors, investors, donors	office furniture and items, technology, software, hardware, equipment and tools	meetings, conference calls, webinars, interviews, conventions, interviews, orientation, fairs, performance reviews	writing, reading, researching, listening, discussing, brainstorming, performing tasks, communicating, setting goals, partaking in meetings	manuals, websites, blogs, resumes, contracts, job descriptions, documents, applications, reviews, minutes, forms

CEFR Descriptors

BSET is aligned to the CEFR descriptors as shown in the following tables:

Overall Descriptors					
A1	**A2**	**B1**	**B2**	**C1**	**C2**
Can understand and use familiar	Can understand sentences and	Can understand the main points	Can understand the main ideas	Can understand a wide range of	Can understand with ease

everyday expressions and very basic phrases aimed at the satisfaction of needs of a concrete type. Can introduce him/herself and others and can ask and answer questions about personal details such as where he/she lives, people he/she knows and things he/she has. Can interact in a simple way provided the other person talks slowly and clearly and is prepared to help.	frequently used expressions related to areas of most immediate relevance (e.g. very basic personal and family information, shopping, local geography, employment). Can communicate in simple and routine tasks requiring a simple and direct exchange of information on familiar and routine matters. Can describe in simple terms aspects of his/her background, immediate environment and matters in areas of immediate need.	of clear standard input on familiar matters regularly encountered in work, school, leisure, etc. Can deal with most situations likely to arise whilst travelling in an area where the language is spoken. Can produce simple connected text on topics which are familiar or of personal interest. Can describe experiences and events, dreams, hopes & ambitions and briefly give reasons and explanations for opinions and plans.	of complex text on both concrete and abstract topics, including technical discussions in his/her field of specialisation. Can interact with a degree of fluency and spontaneity that makes regular interaction with native speakers quite possible without strain for either party. Can produce clear, detailed text on a wide range of subjects and explain a viewpoint on a topical issue giving the advantages and disadvantages of various options.	demanding, longer texts, and recognise implicit meaning. Can express him/herself fluently and spontaneously without much obvious searching for expressions. Can use language flexibly and effectively for social, academic and professional purposes. Can produce clear, well-structured, detailed text on complex subjects, showing controlled use of organisational patterns, connectors and cohesive devices.	virtually everything heard or read. Can summarise information from different spoken and written sources, reconstructing arguments and accounts in a coherent presentation. Can express him/herself spontaneously, very fluently and precisely, differentiating finer shades of meaning even in more complex situations.

Reading Comprehension Descriptors					
A1	**A2**	**B1**	**B2**	**C1**	**C2**
Can understand very short, simple texts a single phrase at a time, picking up familiar names, words and basic phrases and	Can understand short, simple texts containing the highest frequency vocabulary, including a proportion of shared	Can read straightforward factual texts on subjects related to their field of interest with a satisfactory level of comprehension.	Can read with a large degree of independence, adapting style and speed of reading to different texts and purposes, and using	Can understand in detail lengthy, complex texts, whether or not these relate to their own area of speciality, provided they can reread	Can understand virtually all types of texts including abstract, structurally complex, or highly colloquial

rereading as required.	international vocabulary items.		appropriate reference sources selectively. Has a broad active reading vocabulary, but may experience some difficulty with low-frequency idioms.	difficult sections. Can understand a wide variety of texts including literary writings, newspaper or magazine articles, and specialised academic or professional publications, provided there are opportunities for rereading and they have access to reference tools.	literary and non-literary writings. Can understand a wide range of long and complex texts, appreciating subtle distinctions of style and implicit as well as explicit meaning.

Listening Comprehension Descriptors					
A1	**A2**	**B1**	**B2**	**C1**	**C2**
Can understand familiar everyday expressions and very basic phrases about self, family and immediate concrete surroundings when people speak slowly and clearly.	Can understand sentences and frequently used expressions related to the areas of immediate personal relevance. Can catch the main point in short, clear, simple messages and announcements.	Can understand the main points of clear standard speech on familiar matters regularly encountered in work, school, and leisure including short narratives. Can understand factual information about common everyday or job related topics, identifying both general and specific details, provided speech is clearly articulated in a generally familiar accent.	Can understand standard spoken language, live or broadcast, on both familiar and unfamiliar topics normally encountered in personal, social, academic or vocational life. Can understand the main ideas of propositionally and linguistically complex speech on both concrete and abstract topics delivered in a standard dialect. Can follow extended speech	Can follow extended speech even when it is not clearly structured and when relationships are only implied and not signalled explicitly. Can recognise a wide range of idiomatic expressions and colloquialisms, appreciating register shifts. Can understand enough to follow extended speech on abstract and complex topics beyond his/her	Can understand with ease virtually everything heard

			and complex lines of argument provided the topic is reasonably familiar and the direction of the talk is sign-posted by explicit markers.	field.	

Writing Comprehension Descriptors					
A1	**A2**	**B1**	**B2**	**C1**	**C2**
Can ask for or pass on personal details.	Can compose short, simple formulaic notes relating to matters in areas of immediate need.	Can convey information and ideas on abstract as well as concrete topics, check information, and ask about or explain problems with reasonable precision. Can compose personal letters and notes asking for or conveying simple information of immediate relevance, getting across the point they feel to be important.	Can express news and views effectively in writing, and relate to those of others.	Can express themselves with clarity and precision, relating to the addressee flexibly and effectively.	Can express themselves in an appropriate tone and style in virtually any type of formal and informal interaction.

Speaking Comprehension Descriptors					
A1	**A2**	**B1**	**B2**	**C1**	**C2**

Can interact in a simple way but communication is totally dependent on repetition at a slower rate, 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.	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 with some confidence on familiar routine and non-routine matters related to their interests and professional field. Can express thoughts on more abstract, cultural topics such as films, books, music, etc.	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 they want to say, adopting a level of formality appropriate to the circumstances.	Can express themselves 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.	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 that the interlocutor is hardly aware of it.

Test Components

BSET is a comprehensive, computer-based test which measures all four English proficiency skills. It is divided into four components:

Component	Distribution	Items
Listening Comprehension	25%	25
Reading Comprehension	25%	25
Speaking	25%	2
Writing	25%	2

The listening and reading components are composed of selected-response formats and are solely multiple-choice questions. In the speaking and writing component, candidates are prompted to

record their spoken and written answers to tasks. The speaking test is semi-direct as candidates record their answers to pre-recorded questions. There is not an interlocutor. All components and tasks are contextualized within TLU domains and approximate real-life language use situations.

An overview of the structure of each component is included in the appendices.

Listening

Listening comprehension is essential in communication. From the very beginning of language learning, interaction with others requires listening to another speaker and comprehending spoken information. BSET's listening component evaluates the ability of examinees to understand spoken English information.

The BSET Listening component evaluates the examinee's understanding of English in various spoken situations that candidates will likely find themselves in outside of the exam. This section includes 25 multiple-choice questions that require understanding for gist, specific information, main ideas and supporting details, the ability to infer meaning and deduce meaning of target vocabulary, and to recognize attitudes and opinions. The input audios include both monologues and dialogues. Content and items reflect authentic English language use.

The Listening section consists of three parts and is allotted a total of 30 minutes.

Part 1: In this part, candidates listen to short monologues and conversations between two people lasting two to three sentences in length. There are 8 multiple choice questions in this part.

Part 2: In this part, candidates listen to a short conversation between two people. The conversation lasts about 1 to 2 minutes. The conversation is followed by 9 multiple choice questions.

Part 3: In this part, candidates listen to a lecture. The conversation lasts about 4 minutes and is followed by 8 multiple choice questions.

Reading

Being able to read and comprehend written English is important in daily life, in academic settings and in places of employment. Navigating daily life in English requires reading environmental English, studying in English requires reading course materials while working in English requires reading emails, reports and documents. Using context to understand unknown vocabulary, catching main ideas and synthesizing information are all critical for effective reading comprehension and language learning.

The BSET Reading component evaluates the examinee's understanding of written English. It includes 25 multiple-choice questions that require identifying main ideas and details, determining the meaning of unknown words based on context, and synthesizing relevant information.

The Reading component consists of 3 parts and is allotted 40 minutes.

Part 1: In this part, candidates read short advertisements, posts and written communication. There are 8 multiple choice questions in this part.

Part 2: In this part, candidates read a short text of about 250 words. The passage is at an intermediate level and is followed by 9 multiple choice questions.

Part 3: In this part, candidates read a longer text of about 450 words written at an advanced level. It is followed by 8 multiple choice questions.

Writing

Writing in English is a necessary skill for communication, especially in academic and professional environments. The BSET Writing component evaluates the examinee's ability to produce meaningful, appropriate, and accurate pieces of original writing in English.

The Writing component consists of 2 parts and is allotted 30 minutes.

Part 1: In this part, candidates read a short prompt at a low-intermediate level and produce a 50 to 75 word email, note or paragraph

Part 2: In this part, candidates read a statement about a topic and write a response expressing and supporting their opinion. The word count should be between 175 to 225 words.

Speaking

Speaking in English is essential for basic communication in English-speaking environments. Self-identification, making basic requests, asking questions and stating an idea or opinion are all necessary for learners of English. The Speaking component of BSET covers a broad spectrum of knowledge by asking a short personal question and requesting an opinion on a topic.

The BSET Speaking section evaluates the examinee's ability to produce meaningful and comprehensive speech in response to prompts.

The Speaking component consists of 2 parts and is allotted 7 minutes.

Part 1: In this part, candidates are asked a short, personal prompt at a low-intermediate level. Candidates have one minute to plan and up to one minute and a half to record their response.

Part 2: In this part, candidates verbally respond to a given statement. They have two minutes to prepare and up to two minutes and a half to record their opinion.

Section 4: Exam Procedure

Mode of Delivery

BSET is computer-based and is delivered online on the Internet. Computer systems must include headsets and microphones for the speaking and listening components. The test is divided into four sections which are taken together in one sitting. Test items are randomized from a larger item bank to ensure the integrity of the exam. All overall exams are of comparable quality and difficulty. The exam is administered both in-person at certified Bright Online test centers with proctors and in a remote setting from any device using our online proctoring service. In-person exams are arranged together with our partner institutions. Bright Online and partner institutions

agree on designated days and times for exams. Partner institutions gather candidates registration forms and provide a final list of candidates to Bright Online for each exam day. Bright Online then issues login details for the candidates and provides guidelines for test security and technical support during test administration. Partner institutions oversee the exam delivery at certified test centers and follow standardized guidelines for proctoring.

Examinee Experience

Prior to the start of the exam, candidates must present an official government identification either to the exam proctors if in-person or displayed to the computer camera and captured and verified for remote proctoring. Candidates are then guided through a series of compatibility checks. The candidate's computer system is checked for connectivity and functionality of audio playback and audio recording.

After all the checks are completed, candidates can proceed to take the exam. Each of the four sections have a fixed time and are automatically timed within the computer system. Candidates can view the remaining time per section throughout the exam and a progress bar is displayed for visual support. If a candidate finishes a section with time remaining, they can review and revise their answers or they can proceed to the next section. Automatically at the end of the allotted time, candidates will be moved to the next section. During the exam, candidates chose the correct answer from multiple choice options for each question in the listening and reading sections. For the writing section, candidates type their answers directly into a type entry field. Speaking answers are spoken into a headset microphone and recorded in the test-taking system. At the end of the exam, candidates will see a final screen congratulating them on finishing the exam and letting them know that they will be contacted with the results. Candidates receive their results in certificate format directly from us or our partner institutions.

Section 5: Scoring and Proficiency Levels

Test Scoring

An overall English language proficiency score is given and scores are given for each individual language skill, reading, writing, listening and speaking. Scores are allocated based on the

candidate's results in the specific test component. The Listening and Reading components are scored automatically by the testing system. Each question has equal weight and there is no penalty for wrong answers. The Writing and Speaking responses are marked by trained raters. Raters access responses electronically and evaluate the answers according to standardized rubrics. All Speaking and Writing raters have experience in English language teaching and hold recognized English language teaching qualifications. Our professional raters go through training and recalibration every 60 days to ensure consistency and continued adherence to the rubric. Reading, Writing, Listening and Speaking scores are reported on a scale of 100-700. The scale score is not equal to the number of correct items or raw score. Raw scores are converted to scale scores. The overall score is computed from the individual skills and is determined by equal weight given to all four skills: Reading, Writing, Listening and Speaking. Language learners have diverse profiles and their levels may not be the same for each skill; therefore, both the individual skill scores and overall score are reported on an official certificate. The certificates are verifiable online.

Proficiency Levels

Performance on Bright Standard English Test is reported on a scale from 100 to 700.

Table 1: BSET Scores and Levels

100-199:	Breakthrough
200-299:	Waystage
300-399:	Threshold
400- 499:	Vantage
500- 599:	Effective Operational Proficiency
600- 700:	Mastery

Because students of English need to have their English language ability assessed and defined for a variety of reasons, in particular for academic placement, there must be world-wide agreement on how to define the various levels. Commonly used descriptors of English language levels are found in the Council of Europe's Common European Framework of Reference for Languages

(CEFR). Levels determined by CEFR standards are used by institutions such as language schools, colleges and universities, and company language training programs. For this reason, Bright Online has mapped BSET to the CEFR levels.

The correlation between BSET proficiency levels and CEFR is included in the appendices.

Section 6: Test Development

Introduction

The development process of BSET entails four phases or processes: conceptualization, design, production and feedback. The development process is iterative, always adapting and evolving through research, data and reflection.

Figure 1: Overview of Bright Standard English Test process
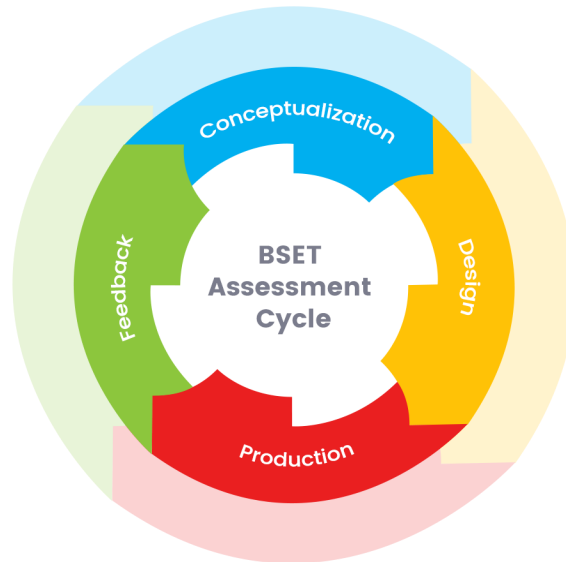
From the onset, Bright Online has used the Common European Framework of Reference, CEFR, as a benchmark for BSET. The CEFR comprehensively defines language proficiency across the main language skills. It divides language proficiency into six levels and details language proficiency descriptors at each level. Bright Online referred to The Council of Europe's Manual for Linking Exams to the CEFR (2009) throughout the test development process specifically

through the stages of familiarization, specification, standardization, standard setting and validation as specified in the manual. Linking BSET to the CEFR levels has been at the forefront of every step of the test development process. Along with rater standardization and internal consistency checks, BSET's alignment to the CEFR levels is strongly validated.

Conceptualization

The concept of Bright Standard English Test stemmed from the need for an accessible and meaningful test of English proficiency suitable for a broad range of stakeholders. During the concept phase, the development team completed a thorough review of current English exams, including a needs analysis. The team further familiarized themselves with the CEFR levels through studying and discussing the Common European Framework. In the final stages, a blueprint for the exam content, structure, delivery and reporting needs was outlined.

Design

After the conceptualization phase, a team of professional English language educators and assessment professionals was formed for the test design phase. This broader team again went through a period of familiarization of the CEFR through discussions and hands-on activities including sorting CEFR descriptors by level. The team then devised appropriate tasks to be used within the exam and wrote detailed test specifications for all tasks and items.

For BSET to be a valid, reliable, fair and accurate measurement of candidates' English language proficiency, tasks are aligned to a CEFR level. Anchor tasks are used to ensure consistency and individual tasks are screened to ensure alignment to the CEFR level. Every item written for and used in BSET measures one of the descriptors described in the CEFR levels. As well as being aligned to the CEFR descriptors and levels, BSET assesses candidates' performance on subskills. Subskills assessed in the listening section include the ability to listen for gist, specific information, main ideas and supporting details, and listening to infer meaning, to deduce the meaning of a word, and to recognize attitudes and opinions. Subskills assessed in the reading section include candidates' ability to read for specific details and main ideas, determine what a

word means based on context and understand why statements are written by connecting relevant information.The specifications for BSET were carefully written and clearly described to enhance not only the quality and reliability of the test, but also to ensure impact and meaning.

An example of a detailed task specification used within the components is included in the appendices.

Production

The goal of the production phase is to produce live versions of Bright Standard English Test. During the production phase, items are written, reviewed, field tested, analyzed and banked to be used in live versions of BSET.

Item writers are native-level speakers of English who have extensive experience in language teaching and assessment. Commissioned item writers go through an internal training process to review key concepts regarding item writing, to provide familiarization with BSET's specifications and item writing guidelines and to further build familiarization and standardization around CEFR. During the training, item writers write sample items and receive guidance and feedback from the trainers. Following the training, item writers receive ongoing mentoring and feedback from quality assurance reviewers.

Quality assurance reviewers manage the entire item writing process from items being commissioned until they are retired. Items are commissioned to be written by the item writers that have successfully passed through our training process. These submitted items are reviewed for content and alignment to CEFR levels by professional reviewers. Reviewers also ensure that there is standardization across the items and tasks by confirming their adherence to the specifications, relation to the anchor items and linked to the correct CEFR level. Additionally, items are reviewed for conformity to testing standards by psychometricians. Each item is scrutinized for content, domain, bias, difficulty and suitability of the stem and choices. Items that do not meet the standards are discarded or sent back to item writers for editing. Items that are

sent back for editing go through another round of review after being resubmitted. Items that meet the standards are passed through to the next step of the process. These passing items are field tested on a select range of students across abilities and contexts. Data from the field test is gathered and analyzed for internal consistency. Items that successfully pass through this rigorous procedure are banked and ultimately used in live versions of BSET. All items used in the exam are retired after a period of time to minimize the possibility of items being compromised by potential sharing between examinees. At any given time, BSET has a bank of approximately 1600 items to be used in live versions of the exam.

Figure 2: Item writing process
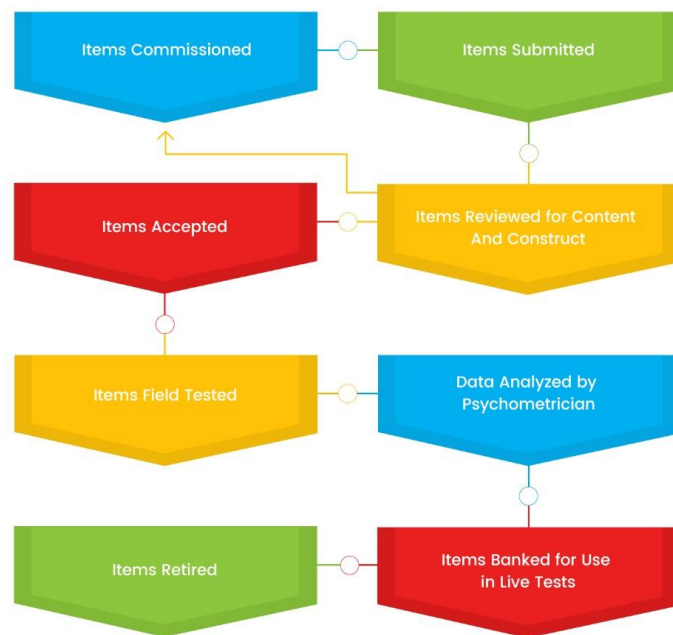
Item Formats

There are two item formats used in Bright Standard English Test; multiple-choice items and constructed-response items.

- **Multiple-choice items** provide valid measures, are feasible to administer and can be scored reliably. They can measure a wide-range of language skills and knowledge in a relatively short testing period. Multiple-choice items are used in the reading and listening sections of BSET.
- **Constructed-response items** allow candidates to demonstrate their productive language learning skills and are used in the speaking and writing sections of BSET.

Item Set Guidelines for Listening and Reading Passages

Item set guidelines are followed for groups of items linked to specific reading and listening passages. Following guidelines ensures exam conformity and unity as well as ensuring a smooth test-taking process for candidates. There are three main guidelines for the writing of item sets:

1. Questions are arranged according to language proficiency levels. Similar to how the passages within the reading and listening sections increase in difficulty, so do the questions. The questions in the beginning correlate to a lower English level than the questions towards the end of the passages and sections. Items become increasingly more difficult as the candidates proceed through the exam.
2. Questions follow the sequential order of the passage. Questions are asked in the same sequence that the answers can be located in the test. This ensures that the candidates are able to answer the questions in the allocated time.
3. Items are independent of each other. Information in the items do not give clues to other passage items. Furthermore, answering a question correctly does not depend on answering any of the previous questions correctly. Each item stands alone.

General Item Guidelines

Guidelines are followed when writing and reviewing test items. The guidelines in this section are important for the quality of test development and effective items. The quality and suitability

of items are judged both by adhering to the specifications and by following the set-forth guidelines.

1. Level appropriateness: The vocabulary, grammar and reading level of the item are aligned to the target CEFR descriptor and level.
2. Testing time: Time required to complete tasks and items are considered when writing items. As an industry standard, each listening question is expected to be answered in one minute, while each reading question is expected to be answered in a minute and a half. This includes the time required to listen to the audio or read the passage
3. Avoiding bias: Items are free from bias. Items were created to not give unfair advantages to groups of candidates. Extra care was given to consider the diverse backgrounds of candidates.

Multiple Choice Guidelines

Specific guidelines for writing multiple-choice items are followed to ensure that items are fair, consistent and reliable The guidelines are specific to writing the stem, structuring the response options and developing distractors.

1. The stem provides enough information to make the task clear and unambiguous. Candidates should be able to answer the question prior to reading the options.
2. The stem is concise. It does not contain extra information.
3. There is only one correct or best answer.
4. All of the response answers are similar in length and complexity.
5. Plausible distractors are used based on likely errors or misconceptions.

Feedback

From the onset, Bright Standard English Test developers realized that large-scale assessments are not static entities. High-stake, large-scale assessments should evolve according to ongoing research, data analysis and reflection. In this regard, Bright Standard English Test has been designed to be dynamic. Routine feedback is received from all stakeholders, including

test-takers, test administrators and partners. Live test data is compiled from a broad range of test dates and centers. Furthermore, BSET professionals are actively involved in international testing associations and stay up-to-date on assessment research and methodology. Gathering this wealth of feedback is influential in preparing for new versions of the assessment. Through the feedback stage, areas of strengths and areas for further refinement and innovation are identified.

Section 7: Reliability, Validity, Authenticity and Washback

Introduction

In order for tests to be substantial, accurate, and flawless, they must be free of bias and distortion. *Bias* is a strong belief that something is better, or more correct, based on opinion and not on fact; *distortion* is the act of taking a bias and presenting it as if it were fact. There are many considerations regarding language including background knowledge and experience of the cultural nuances of the target language which second language learners often do not bring to their language acquisition process. For this reason, it is imperative that language assessment tools do not blur these kinds of information in the testing process. Furthermore, BSET is fully in alignment of best practices in current language teaching and learning, and best practices in methodology, focused on the Communicative Approach.

In order to avoid bias and distortion, BSET is designed within the constructs of reliability, validity, authenticity, and washback.

For the purposes of describing the reliability, validity, authenticity and washback of BSET, the following definitions of these terms are used:

- Reliability: consistency in the measurements regardless of other outside factors. For example, a reliable vehicle starts regardless of the time of day, weather, or temperature.
- Validity: accuracy of exam questions to test the intended target. For example, a language assessment that uses cultural knowledge to test knowledge of grammatical structure is not necessarily valid.

- Authenticity: replication of real-work language use. For example, authentic assessment measures how well the examinee can understand and appropriately respond in real situations outside the classroom.
- Washback: exam effects on teaching and learning. For example, a test that focuses too heavily on how a language works instead of language *use* would create adverse teaching and learning environments where students learn more *about* a language rather than how to use it.

Reliability

Reliability refers to the consistency of the exam so that it produces the same results no matter who the examinees are. To ensure reliability, BSET underwent the following standard measures:

1. Test-retest reliability measures the consistency of results when the same test is repeated on the same sample at a different point in time. BSET was administered twice to the same group of 689 candidates with 14 days between test administrations. Test-retest values can range from -1.0 to 1.0. Values exceeding .70 are considered reliable. The overall test-retest reliability was .81.

2. Internal consistency measures the correlation between multiple items in a test that are meant to measure the same construct. Measurements of internal consistency require that the assessment component contains multiple items, because of this internal consistency is only appropriate for the listening and reading components. The Kuder-Richardson Formula 20 (KR20) was used to compute internal consistency of the listening and reading components of Bright Exam. Values between .70 and .90 are acceptable. The KR20 for Listening and Reading was .76 and .79, respectively.

Validity

The term validity refers to how well an assessment actually measures what is intended to be measured. BSET was developed and designed to provide English language proficiency scores for a broad range of test-takers and stakeholders from academic to professional settings. There are various ways to measure validity, one of which is content validity or content validation.

Content validation ensures that the content of an assessment matches the real-world knowledge, skills, and aptitudes it is designed to measure. Evidence of content validation can be both logical and empirical in nature and can involve the analysis of items, delivery methods and scoring. Evidence of content validity in BSET include the following:

- Task topics are related to actual English use in public, personal, educational and occupational TLU domains.
- Content and item writers are trained experts in the field of English language learning and assessment. Items are reviewed for cultural bias. Items do not reflect specialized or common knowledge.
- Content and scoring are mapped to the CEFR.

Authenticity

The authenticity component of a language assessment must determine how well the examinee can function in a new situation that is happening in real time with fluid language input and requests of language production. In other words, the language in the exam must not be static, old, unused language that is seldom heard in the real world, rather it should be current and actively a part of daily communications. BSET materials were collected from authentic sources such as news clips, podcasts, print materials in the current market and real-time conversations.

Washback

Washback refers to how testing and examinations affect how teachers teach and what learners learn. BSET directly evaluates candidates' proficiency levels in the four skill areas of reading, writing, speaking and listening. Grammar and vocabulary are not isolated factors, but embedded in communicative tasks throughout the content and skill areas of the exam. The focus on real-world use of English in the four skill areas promotes meaningful, authentic and communicative language teaching and learning where students learn to use English.

References

Bachman, L. and Palmer, A. (2010) Language Assessment in Practice. Oxford: Oxford University Press.

Chapelle, C. A. (2008). "Utilizing technology in language assessment," in *Encyclopedia of Language and Education, 2nd Edition, Volume 7: Language Testing and Assessment*, eds E. Shohamy and N. H. Hornberger (Boston, MA: Springer), 123–134.

Council of Europe. (2001). Common European Framework of Reference for Languages. Cambridge: Cambridge University Press.

Council of Europe. (2018). Common European Framework of Reference for Languages: Learning, teaching, assessment; Companion volume with new descriptors. Retrieved from www.coe.int/lang-cefr

Powers, Donald E. (2010). "The Case for a Comprehensive, Four-Skills Assessment of English Language Proficiency," TOEIC,

Appendix A: Overview of the Structure of BSET

Overview of the structure of BSET listening component

Part	Items	Format	Task description	Response format
1	8	Monologues & Dialogues	Listen to short monologues and conversations to identify specific details and main ideas.	4-option multiple choice. Some lexical overlap between distractors and words in the input text.
2	9	Dialogue	Listen to one two-to-three minute intermediate level conversation to identify both concrete and abstract information.	4-option multiple choice. Target and distractors are paraphrased when possible. Distractors refer to topic-related information and concepts mentioned in the text, but are not possible answers.
3	8	Monologue	Listen to one four-minute lecture at an advanced level to identify the speaker's attitude and opinion. The targeted information requires integration of different sections of the text to solve the correct answers.	4-option multiple choice. Target and distractors are paraphrased. Distractors refer to information and concepts mentioned in the text, but are not possible answers.

Overview of the structure of BSET writing component

Part	Time given	Task description	Expected output	Rating criteria
1	5 minutes	Read a short prompt at a low-intermediate level and write an appropriate response	50-75 word note or paragraph	Rating criteria is based on the following aspects: 1. grammatical range and accuracy 2. vocabulary range and accuracy
2	25 minutes	Read an advanced level topic statement and write a response expressing and supporting an opinion	175-225 word text	3. punctuation 4. task fulfillment/topic relevance 5. cohesion

Overview of the structure of BSET speaking component

Part	Time to plan	Time for reponse	Task description	Rating criteria
1	1 minute	1.5 minutes	Read a short question at a low-intermediate level. Prepare and record an answer to the question.	Rating criteria is based on the following aspects: 1. grammatical range and accuracy 2. vocabulary range and accuracy 3. pronunication 4. fluency 5. task fulfillment/topic relevance
2	2 minutes	2.5 minutes	Read a short advanced level statement on an issue. Prepare and record a response expressing an opinion on the topic.	

Overview of the structure of BSET reading component

Part	Items	Task focus	Task description	Response format
1	8	Word and sentence level meaning	Read advertisements, posts and written communication at a beginner level. Requires word and sentence-level comprehension.	4-option multiple choice. Some lexical overlap between distractors and words in the input text.
2	9	Text-level comprehension of short text	Read one intermediate level passage, about 250 words in length. Requires text-level comprehension and reading across multiple sentences.	4-option multiple choice. Target and distractors are paraphrased when possible. Distractors refer to topic-related information and concepts mentioned in the text, but are not possible answers.
3	8	Text-level comprehension of longer text	Read one advanced level passage, about 450 words in length. Requires comprehension of complex and abstract ideas.	4-option multiple choice. Target and distractors are paraphrased. Distractors refer to information and concepts mentioned in the text, but are not possible answers.

Appendix B: Correlation between BSET and the CEFR

BSET	CEFR	Reading	Writing	Listening	Speaking
600 - 700	C2 Mastery	• Comprehends virtually all aspects of a wide variety of academic material for nonspecialists • Rarely requires use of dictionary	• Writes complex documents such as research reports using appropriate style and vocabulary	• Comprehends virtually all spoken language • Understands English spoken in a variety of non-native accents	• Communicates accurately and effectively on practically all academic and social topics in culturally appropriate ways
500 - 599	C1 Advanced	• Understands main ideas and most of the details of factual and literary texts	• Writes clear, well-structured text, expressing points at some length • Writes about complex subjects in a letter, essay or report	• Identifies attitude and purpose of speakers • Grasps main ideas and the majority of supporting details from lectures	• Interacts fluently and spontaneously without much searching for expressions • Formulates ideas and opinions with precision and communicates skillfully with others
400 - 499	B2 Upper Intermediate	• Gathers most main ideas from articles and reports, but has an uneven grasp of details • Interprets meaning of complex sentences and new vocabulary using clues	• Writes reasonably coherent informative and opinion essays on familiar topics • Writes letters about personal significance of events and experiences	• Identifies main ideas and details in conversation • Begins to determine the attitudes of speakers • Understands main ideas from extended speech and lectures on familiar topics, but misses significant details	• Interacts with a degree of fluency and spontaneity • Begins to express abstract concepts, especially on familiar topics • Explains viewpoint on issues, giving disadvantages and advantages
300 - 399	B1 Intermediate	• Understands the description of events, feelings and wishes • Comprehends familiar subjects, and identifies some significant details	• Writes simple connected texts on familiar topics or topics of personal interest. • Describes experiences and impressions in personal letters	• Understands the main point and the general outline of standard speech at work, school or leisure	• Manages day-to-day communications marked by frequent grammar and vocabulary errors • Gives reasons and explanations for opinions and plans
200 - 299	A2 Elementary	• Finds predictable information in simple material • Comprehends short simple personal letters • Begins to determine the meaning of words by context	• Writes short, simple notes, messages, and personal letters relating to immediate needs • Uses only basic vocabulary and simple grammatical structures	• Comprehends the main point in short, clear messages and announcements • Follows a simple conversation on familiar topics • Understands very basic exchanges when spoken slowly using simple vocabulary	• Communicates in simple and routine tasks with simple questions, greetings, expressions of needs, and preferences • Uses phrases and sentences to describe family, people, lliving conditions, educational background and career
100 - 199	A1 Beginner	• Comprehends familiar words and very simple sentences • Demonstrates understanding of a few simple grammatical and lexical structures	• Writes only short, simple sentences • Provides personal details with correct spelling and can copy familiar words and phrases	• Understands simple greetings, statements, and questions when spoken slowly and clearly • Follows simple familiar instructions • Understands a few isolated words or phrases spoken slowly	• Interacts in a simple way with repetition and repharsing • Asks and answers simple questions on very familiar topics • Uses simple phrases and sentences to describe familiar places or people

Appendix C: Sample Specifications

Listening: Part 2					
Task					
Skill	Understanding discourse and meaning in abstract texts				
CEFR level	A1	A2	B1	B2	C1	C2
Description	Candidates listen to a dialogue between two people. This task is most likely encountered in educational a				
Instructions	Listen to the following conversation and answer the following 9 questions				
Response Format	4-option multiple-choice questions				
Time	30 minutes for the entire listening component				
Input					
Length	2 to 3 minutes				
Domain	Public	Occupational	Educational	Personal	
Discourse mode	Descriptive	Narrative	Expository	Argumentative	Instructive
Content knowledge	General		Specific		
Cultural specificity	Neutral		Specific		
Nature of information	Only concrete	Mostly concrete	Fairly abstract	Mainly abstract	
Presentation	Verbal	Non-verbal	Both		
CEFR level	A1	A2	B1	B2	C1	C2
Grammatical level	A1-B2 (See Grammatical Level Guidelines)				
Avg. sentence length	13 to 20 words				
Topic	From B1 and B2 topic list				
Input Types	recordings of dialogues				
Input Sources	radio, phone, internet, audio books, self-created, films, TV, advertisements, announcements, lectures, classroom discourse				
Authenticity	As authentic and non-scripted as possible				
Speaker Characteristics	male or female, any age, any accent, 2 voices maximum				
Response					
Question Types	Mulitple choice with 3 distractors and 1 correct response				
Distractors	Must be viable and refer to topic-related information and concepts mentioned in the text Paraphrased				
Targeted Information	Across sentences and paragraphs May be implied Paraphrased				
Grammatical level	A1-B1				