

Original Paper

Validating an EFL Summative Test in Business English Major at a Chinese University

Rong Zhang¹

¹ Rong Zhang, Institute of Education, University of Reading, Reading, United Kingdom

Received: June 22, 2023

Accepted: July 29, 2023

Online Published: August 05, 2023

doi:10.22158/eltls.v5n3p167

URL: <http://dx.doi.org/10.22158/eltls.v5n3p167>

Abstract

The research did an analysis on content validity, construct validity and concurrent validity of an summative assessment for an Advanced Business Course administered by a university in Guangdong University, China. With validity-based assessment, this study provided implications for educators and practitioners on improving quality of test papers.

Keywords

content validity, construct validity, concurrent validity, summative assessment

1. Introduction

1.1 An Overview of Business English

The arrival of digital information era has driven the need for versatile talents. As companies are on the track of digital transformation, employees are expected to develop a wide repertoire of skills, particularly background knowledge about trade, business and foreign languages to meet the diverse requirements of workforce. English remains to be the most frequently used language in business context. Under such circumstances, Business English has altered from an interdisciplinary program to an independent major in Chinese tertiary education. In this section, I am going to briefly recall significant changes of Business English in Asian context.

The origin of Business English major can be traced back to professional training of translation about foreign trade in 1950s. Initially, textbooks were used on class to inform students of basic principles in real-context business transaction and manners required in the profession, including: “making enquiries, offer and counter-offer, translation of business terms and letters”. In 1980s, Reform and Opening-up Policy indicates the need to train and cultivate business leaders with proficiency in English. In response, new academic programs were introduced into Chinese Universities, namely marketing, international relations and management, whose approach is known as English as a Medium of Instruction (EMI). It

was not until 2007 that Business English has been officially approved by Chinese Education Ministry as a major, marked by an attempt to run a pilot scheme at University of International Business and Economics (UIBE).

1.2 Business English AND English for Specific Purpose (ESP)

It has been widely assumed that Business English lies in the field of English for Specific Purposes (ESP) (see Dudley-Evans and St Johns, 1998, as cited in Bargiela-Chiappini & Zhang, 2012). There are a wide range of practices integrating Business English into the curriculum of English for Specific Purposes (ESP), despite variety in linguistic and social context. In Asian countries, unsurprisingly, Business English “brand” continues to be strongly influenced by “Western preferences”, particularly the theories and methods of international business and crosscultural management to be taught alongside or as components of business communication programs (Bargiela-Chiappini & Zhang, p. 194).

Initially, Zhang (2005, 2007) proposed a discourse approach to Business English, arguing that language should be practiced and taught within business activities and contexts. On top of that, a pedagogical approach to Business English was proposed (Zhang, 2012), which involves “the teaching of system of strategic communication in international business in which participants adopt business conventions and procedures and make selective use of lexico-grammatical resources of English as well as visual and audio semiotic resources to achieve their communicative goals” (Zhang, 2012, p. 48). Instead of insisting on language teaching, the latter put more emphasis on business activities in real-world context in which language is only a tool to convey insightful ideas for communication purposes.

Despite emergent perspectives and approaches to Business English, terms are often too technical for teachers who are neither specialists nor researchers. Worse still, in China, Business English in traditional class is of boredom, disdain and somewhat remote from reality. To illustrate, according to Zhang (2012), a survey of Business English textbooks from the UK and the US (Nelson, 2000, 2006, cited in Nickerson, 2010) found that their contents bore little relevance to authentic workplace written and spoken English.

Although the discrepancy between teaching resources and actual work practices seems to be a problematic issue, what is more disconcerting is that shortcomings pointed out in the literature as early as the 1980s should continue to be propagated by influential publications exported all over the world. Specifically in China, imported materials are directly in use on class, without being explored whether they can be fully adapted to local domestic context. The textbooks currently in use disconnect from real-world context, and examination papers are mainly based on memorized knowledge and rote learning outcomes. Curriculum design of Business English also remains unchanged throughout the past decade. According to Bargiela-Chiappini and Zhang (2012), “ESP is an essentially materials-driven and teaching-led movement” (Bargiela-Chiappini & Zhang, 2012, p. 149). Therefore, the direct use without major modifications would cause serious problems to development of Business English curriculum.

Several studies have identified the problems with curriculum evaluation in Business English Majors. For example, according to Wen (2022), based on Internet of Things Business Model, the current curriculum system for business English majors has some inherent problems: curriculum objectives, implementation and evaluation. In response, this research is going to analyze a test paper of a course to identify problems in Business English Course and provide implications to practitioners.

1.3 Validation in Assessment

Validity is a central concept in assessment and measurement (Lane et al., 2015; Green, 2014; Bachman & Palmer, 2010). Validity indicates whether we have theoretical and empirical evidence to support inferences we have made based on test scores. Validity has been widely assumed to be one of the most important cardinal criterion for “testing a test” (Brown, 2004) by providing accurate measures of precisely the abilities in which we are interested. Moreover, it has been included as a key element that contributes to the framework of test validation (Bachman & Palmer, 2010). It is argued that test design begins with collection of validity evidence and culminates in intended interpretation of test scores for specified uses.

Domestic studies exploring validity evaluation are mostly aimed at measuring high-stakes assessment, such as College Entrance Examination and standardized language tests, e.g., TOEFL and IELTS. Very little empirical research has been done on validating a summative assessment of a Business English course. In response to the research gap identified, the present study intends to conduct textual analysis evaluating content validity, construct validity and concurrent validity of a summative assessment for an Advanced Business English Course administered by a university in Guangdong province. With validity-based measurement, this study will provide implications to teachers in improving the quality of test papers.

2. Literature Review

Previous studies have highlighted the significance of validity for educators and administrators to incorporate and interpret the scores of students. In other words, validation is crucial for a critical analysis of the approaches in which tests are used to assure the fairness and ethics (Bachman, 2000). Specifically, if teachers can evaluate formative assessment by means of validity theory, they could modify their own teaching methods and practices to help students succeed at language learning tasks so that students would be engaged (Liosa, 2011). By contrast, an assessment that does not test what is being taught gives unfair grades to students and provides negative washback.

Given the diversity in test-takers and assessment purposes, there is no one-size-fits-all solution to measuring validity of an assessment (Douglas, 2010). Validity has long been regarded as an indispensable component of a well-designed assessment (Bachman & Palmer, 2010). If a test measures what it purports to measure, then it is considered to be valid. However, based on another definition given by American Educational Research Association et al. (1999) which highlights the importance of “evidence” and “theory” in justifying the validity, Green (2014) stated that it is important to note that

validity is not a quality of an assessment, but a quality of inferences made upon the results. From my own perspective, the newly-constructed definitions (e.g., Green, 2014) describes the matching degree of interpretations of assessment results and the original purposes of assessment. The only difference is that the methods to collect evidence are getting diverse.

To establish the validity of an assessment, a few questions need to be addressed: “Why do we measure validity of an argument?” and “How can we measure validity?” It is from both internal and external perspectives that validity can be established. Internal validity refers to face validity and content validity, while external validity refers to concurrent validity and predictive validity. Furthermore, Messick (1989) has proposed a unitary approach to validate, i.e. construct validity. In order to get a full picture of validity, the present study explores content validity, concurrent validity and construct validity, so that the assessment can be internally, externally and uniformly validated.

2.1 Content Validity

Validation refers to the process of collecting evidence in order to evaluate the validity of a given test. According to Bachman and Palmer (1996), validity is one of the six elements that comprise test usefulness. According to Haynes et al. (1995), content validity refers to “the extent to which a test adequately and sufficiently measures the particular skills or behavior that it is set out to measure”. When analyzing the content validity of achievement tests, it is advised to compare test specification and test content and to justify if test items cover requirements of test syllabuses. Since the syllabus of a summative assessment is largely based upon the course syllabus, the present study will compare test content and course syllabus of Advanced Business English Course.

2.2 Construct Validity

Construct validity demonstrates the association between test scores and theoretical variables. Based on claims proposed by Douglas (2010), construct validity asks “Does this test actually tap into the theoretical construct as it has been defined”. It examines whether a test covers the full range of behaviors that make up the construct being measured. Accordingly, tests are defined as “operational definitions of constructs in that they operationalize the entity that is being measured.” (Douglas, 2010) Construct validity relates to the accuracy of an assessment, i.e., if the assessment sufficiently evaluates all essential “important concepts, attributes, qualities or traits, e.g., mathematical reasoning, reading comprehension, intelligence or creativity”. Messick (1989) proposes two key terms, namely “construct irrelevance” and “construct under-representation” as two threats of construct validity. Construct irrelevance occurs when elements which test-developers intend to measure are not related to that required in learning outcomes. Factors that influence student test scores but are not directly related to the construct may be called as “construct-irrelevant variable”. In contrast, construct under-representation occurs when some important constituents are missed out. For example, considering the cost of test-related equipment and facilities, speaking tests are often excluded. And results of such proficiency tests may not reflect the overall proficiency level and may fail to give implications about how much progress the student has made in terms of speaking.

2.3 Concurrent Validity

Concurrent validity, along with “predictive validity”, lies in the field of criterion validity. Criterion-related validation derives from comparison between an assessment and another “criterion measure” (Green, 2014, p. 79). While predictive validity indicates the ability of an assessment to make prediction on outcome variables, concurrent validity refers to the extent of agreement between a new assessment and a well-established one which has proven to be valid (McKim, 2022; Cohen et al., 1996). This research adopts TEM-4 (Test for English Majors, Band 4) as the “criterion assessment”, which has been well-established over the past two decades nationwide.

Thus far, the literature review section has discussed three kinds of validity. Turning now to previous studies that are associated with assessment and evaluation. The greater part of the literature is extensive and focuses particularly on K-12 learners (Yumsek, 2023; Wolf, 2022; Faulkner-Bond et al., 2018). In addition, research mainly focuses on standardized test (Ihlenfeldt & Rios, 2023) and lacks clarity regarding a particular summative assessment whose purpose is to evaluate the performance of students at the end of a course.

In response, this research is going to check the test scores of a particular summative assessment against teaching materials and syllabus so as to evaluate the three kinds of validity as mentioned above.

The present study addresses the following questions:

Firstly, to what extent is the assessment valid in content?

Secondly, to what extent is the assessment valid in construct?

Thirdly, to what extent does this assessment correspond to the results of TEM-4?

3. Method

3.1 Participants

Participants were selected from Grade 2 of Business English major at the same university in Guangdong Province, China. Participants' ages ranged from 20 to 23 years ($M=22$, $SD=2.4$) with 43% identified as female and 57% as male. All of the participants were English as Foreign Language (EFL) learners who picked up English approximately at the age of 8 or 9, and they took the summative test specifically designed for Advanced Business English Course.

3.2 Instruments

The present study adopted a qualitative textual analysis of a test paper (content, objectives and teaching syllabus), along with statistic description of test scores (including but not limited to: item types, quantity and weight).

3.3 Materials

This test paper, teaching syllabus specifically designed for Business English major are attached to the appendix section.

4. Result

4.1 Teaching Content Analysis

Integrated Course of Business English (3rd) is a compulsory course specifically designed for third-year college students who major in Business English. The textbook used for this course is Book II within An Integrated Course of Advanced Business English series. The textbook series include Advanced Comprehensive Business English, Advanced Practical Business Writing in English, Up-to-date Correspondence for Import and Export, Advanced Business English Interpreting, etc. The test in the present study is based on Book II, which involves economy, management, investment and market-related knowledge, covering a wide range of topics, e.g., business ethics, market expansion, financial investment, business management, human resources, division of labor, decision making and business negotiation.

The textbook is composed of 12 units. Each unit is made up of text A, text B and supplementary exercises. The exercises mainly include reading comprehension, cloze, paraphrasing, translation, listening and writing. With a wide range of knowledge and sufficiently enough exercises, the textbook is a reasonably good option for Business English major students.

4.2 Syllabus

Syllabus describes what content a course is going to cover, what abilities it is aimed at cultivating and how this course fits into the talent training program of a particular major. A language teaching syllabus “reflects different philosophies of language learning” (Winke & Brunfaut, 2021).

The assessment this study aims at validating is a summative test, which is designed to evaluate students’ learning outcomes over a whole semester. Thus, test syllabus largely relies on teaching syllabus. The syllabus of Integrated Course of Business English (3rd) clearly states that after a whole semester, student’s ability to listen, speak, read, write and translate should be enhanced. They should master the knowledge of pronunciation, vocabulary, grammar, discourse and pragmatics. Apart from general knowledge required for English majors, Business English majors also need to have a rough idea of economics and management, as well as international commercial law, international finance, financial management, etc. And they ought to use this kind of common sense to make judgments where necessary. Besides, the syllabus clearly states the specific tasks for this semester. The course covers unit 7 through unit 10, with each unit taking up 8 class hours.

4.3 Content Validity

The assessment is composed of five sections. Score, weight and source of items in each section are listed in the Table below.

Table 1. Score, Weight and Source of Items

	Item type	Score(points)	Weight	Reference
Section 1	Blank-filling	10	10%	Textbook Unit7-9
Section 2	Paraphrasing	20	20%	Textbook Unit7-9
Section 3	Proofreading	10	10%	TEM-8 Error correction
Section 4	Reading comprehension	30	30%	Foreign magazines
Section 5	Translation	30	30%	Government report of work

The first section asks students to pick up words from the given box and complete the sentences. Test-takers can change the forms if necessary. There are 10 words and 10 phrases in total. Out of the 10 words, 3 words are derived from unit7, 4 words from unit 8, 3 words from unit 9. According to teaching syllabus, this semester covers unit7 through unit 10, words from unit 10 are missed out. What's more, all words and blank-filling items are copied from original exercises in the textbook without essential adjustments.

The second section is paraphrasing. Students are required to rewrite sentences in their own words with little prompt information (e.g., linking words) given. As with section I, so with section II. Sentences to be re-written all come from original textbooks, without any adaptation or revision.

In terms of Section 3, proofreading items are adapted from one of the practice tests for TEM-8. Reading comprehension comprises the third section. There are 3 reading passages, each is composed of approximately 800 words along with 4 multiple-choice questions (2 points each) and one short-answer question (1 point each). Words and phrases are within the Teaching requirements for business English Majors in Colleges and Universities. Number of words and readability-related statistics are listed below.

Table 2. Readability of each Passage in Section 4

Passage Num.	Words	Flesch Reading Ease	Flesch-Kincade Grade Level
No.1	637	52.2	10.9
No.2	828	51.1	10.9
No.3	834	44.4	12.3

The scores of three passages range from 0~60, are typically regarded as college or graduate level. To be more specific, passage one and two are difficult to read, passage three is difficult to read.

Flesch-Kincaid Grade Level indicates the readability based on U.S. school grade levels. It is universally acknowledged that most professional documents fall in the range of 7.0~11.0. In conclusion, the three passages basically comply with the requirements for English major students.

The theme of translation items is “environment-friendly policy”, related to Unit7 “clean energy”. Based on Syllabus of Test for English Majors Band 8, translation section is a Chinese-to-English translation item, approximately 150 words. Therefore, translation section basically meets the requirements of Test for English Majors Band 8. However, according to Syllabus of Test for Business English Majors Band 4, to complete translation items requires general knowledge of international business. Therefore, it is advised that test-developers select another essay with about 150 Chinese characters, whose content involves economics, commerce, management, law, etc. In conclusion, 30% of test is textbook-based, while 70% is syllabus-based. Inferences made upon this test are valid. Furthermore, to examine students’ mastery of knowledge on textbook, it is advised to focus on the first two parts. Remaining three sections are better for making inferences about whether students possess the ability required on the teaching and test syllabus.

In conclusion, based on contrast analysis of test content and course objectives along with specifications, this summative test is an adequate sample which involves ensuring all major aspects are covered and they are in suitable proportions.

4.4 Concurrent Validity

There is no agreement among scholars in choosing the existing criterion measure when seeking concurrent validity. According to Miller et al. (2011), concurrent validity is justified based on the degree in which the results of a test highly corresponds with an assessment that has gone through validation process before. However, some scholars refer to normal records (i.e., results of mid-term tests or formative tests) as the measure and run correlation analysis to see if there is any consistency between performance at ordinary times and summative assessment.

From my point of view, both approaches should be taken with caution. Firstly, criterion measure is supposed to share common constructs with the assessment whose validity is to be certified, i.e., to compare results of a junior-high-school math test with IELTS makes no sense in validation. Secondly, normal records can be used as a reference but not the determinant of concurrent validity only if their validity is established before. Therefore, the present study uses TEM-4, a national English proficiency test that every second-year English major students are supposed to take part in, as a criterion measure. Based on Miller et al.(2011), this study runs correlation analysis and linear regression to validate the assessment.

Table 3. SPSS Results of Normality Test

Test of Normality						
	K-S test*			S-W test		
	Statistic	df	Sig.	Statistic	df	Sig.
Total	.151	53	.004	.934	53	.006
TEM4	.120	53	.056	.964	53	.111

* Lilliefors Significance Correction.

According to the SPSS normality test results total score is non-normally distributed ($p < .05$) while TEM4score is normally distributed ($p > .05$). Thus, both Pearson and Spearman correlation analysis need to be demonstrated as follows.

Table 4. Relationship between Total Score of the Summative Test and TEM-4 Score

	N	r	rho
Total score & TEM-4score	53	.750**	.695*

** $p < .01$.

There is a strong, significant relationship ($\rho = .695$, $r = .750$, $p = .01$) between total score and TEM-4 score for all participants. Thus, the summative assessment and TEM-4 function similarly and they have concurrent validity. Suppose the normal records are validated, let's see if the summative assessment and formative assessment have concurrent validity.

Table 5. Relationship between Final Results and Operating Results

	N	rho
Final results & Operating results	53	.728*

** $p < .01$.

There is a strong, significant relationship ($\rho = .728$, $p = .01$) between final results and operating results for all participants. Thus, the summative assessment and formative assessment similarly and they have concurrent validity, i.e., students performance is similar either at ordinary times or at final exam.

5. Discussion

The present study validated the an EFL summative test specifically designed for third-year university students who major in Business English. It is certified that the assessment has reasonably good content validity, construct validity and concurrent validity (in reference to TEM-4 results).

The main limitation of this study is that the researcher did not get access to sufficiently enough data for students' score of each question but only the total score of a particular part. Or else SPSS reports, primary component analysis in particular might be more accurate. Besides, the present study did not gather data about student feedback after participating the test, since students' comments might be used as useful indirect qualitative evidence for validation process.

The present study fills in the empirical gap of summative assessment validation. It provides pedagogical implications to test-developers that in order to develop high-quality test papers, they should gather comprehensive evidence about quality of papers, especially the validity. And the study also sets an example for future studies exploring validity-related issues of a particular summative assessment.

Acknowledgement

I want to express my greatest gratitude to Mr. Liao Huan, Eric and Dr. Kari Sahan, my supervisors at University of Reading for providing tons of suggestions and answering with unfailing patience numerous questions.

References

- Alderson, J. C., Clapham, C., & Wall, D. (1995). *Language test construction and evaluation*. Cambridge University Press.
- American Educational Research Association 1999 Annual Meeting April 19-23, 1999, Montreal, Canada. (1998). *Educational Researcher*, 27(8), 49-56. <https://doi.org/10.3102/0013189X027008049>
- Bachman, L. (2000). Modern language testing at the turn of the century: Assuring that what we count counts. *Language testing*, 17(1), 1-42. <https://doi.org/10.1177/026553220001700101>
- Bachman, L. F., & Palmer, A. (2010). *Language Assessment in Practice: Developing Language Assessments and Justifying their Use in Real World*. Oxford University Press.
- Bargiela-Chiappini, F., & Zhang, Z. (2012). Business English. In B. Paltridge, & S. Starfield (2012), *The Handbook of English for Specific Purposes*. <https://doi.org/10.1002/9781118339855.ch10>
- Brown, H. D. (2004). *Language assessment: Principles and classroom practices*. Pearson Education.
- Cohen, R. J., Swerdlik, M. E., & Phillips, S. M. (1996). *Psychological testing and assessment: An introduction to tests and measurement* (3rd ed.). Mayfield Publishing Co.
- Douglas, D. (2010). *Understanding language testing*. Routledge.
- Faulkner-Bond, M., Wolf, M. K., Wells, C. S., & Sireci, S. G. (2018). Exploring the factor structure of a

- K-12 English language proficiency assessment. *Language Assessment Quarterly*, 15(2), 130-149.
<https://doi.org/10.1080/15434303.2017.1419247>
- Green, A. (2014). *Exploring Language Assessment and Testing: Language in Action*. Routledge.
<https://doi.org/10.4324/9781315889627>
- Haynes, S. N., Richard, D. C. S., & Kubany, E. S. (1995). Content validity in psychological assessment: A functional approach to concepts and methods. *Psychological Assessment*, 7(3), 238-247.
<https://doi.org/10.1037/1040-3590.7.3.238>
- Ihlenfeldt, S. D., & Rios, J. A. (2023). A meta-analysis on the predictive validity of english language proficiency assessments for college admissions. *Language Testing*, 40(2), 276-299.
<https://doi.org/10.1177/02655322221112364>
- Lane, S., Raymond, M. R., & Haladyna, T. M. (Eds.). (2015). *Handbook of Test Development* (2nd ed.). Routledge. <https://doi.org/10.4324/9780203102961>
- Llosa, L. (2011). Standards-based classroom assessments of English proficiency: A review of issues, current developments, and future directions for research. *Language testing*, 28(3), 367-382.
<https://doi.org/10.1177/0265532211404188>
- McKim, C. (2022). *Validity: Criterion, Concurrent, Ecological, and Predictive*. Routledge.
<https://doi.org/10.4324/9780367198459-REPRW156-1>
- Messick, S. (1989). Validity. In R. L. Linn (1989). *Educational measurement* (pp. 13-103). Macmillan Publishing Co.
- Miller, L. A., McIntire, S. A., & Lovler, R. L. (2011). *Foundations of psychological testing: A practical problem* (3rd ed.). Sage Publications, Inc.
- Nelson, M. (2000). *A corpus - based study of Business English and Business English teaching materials* (Unpublished PhD thesis). Manchester, UK: University of Manchester.
- Nelson, M. (2006). Semantic associations in Business English: A corpus-based analysis. *English for Specific Purposes*, 25. <https://doi.org/10.1016/j.esp.2005.02.008>
- Nickerson, C. (2010). The Englishes of business. In A. Kirkpatrick (Ed.), *The Routledge Handbook of World Englishes*. Routledge .
- Rix, S. (2012). Construct Validity in Formative Assessment: Purpose and Practices. *Journal on English Language Teaching*, 2, 1-6. <https://doi.org/10.26634/jelt.2.2.1823>
- Wen, P. (2022). The curriculum system of business english majors in higher vocational colleges from the perspective of the internet of things business model. *Wireless Communications & Mobile Computing*. <https://doi.org/10.1155/2022/6243729>
- Winke, P., & T. Brunfaut (2021). *The Routledge Handbook of Second Language Acquisition and Language Testing*. Routledge. <https://doi.org/10.4324/9781351034784>
- Wolf, M. K. (2022). Interconnection between constructs and consequences: A key validity consideration in K-12 english language proficiency assessments. *Language Testing in Asia*, 12(1).
<https://doi.org/10.1186/s40468-022-00194-1>

- Yumsek, M. (2023). Educational L2 constructs and diagnostic measurement. *Language Testing in Asia*, 13(1). <https://doi.org/10.1186/s40468-022-00214-0>
- Zhang, Z. (2005). *Business English: A Discourse Approach*. Beijing: University of International Business and Economics Press.
- Zhang, Z. (2007) Towards an integrated approach to teaching Business English: A Chinese experience. *English for Specific Purposes*, 26, 399-410. <https://doi.org/10.1016/j.esp.2006.10.006>
- Zhang, Z. (2012) Teaching Business English. In C. A. Chapelle (Ed.), *The Encyclopedia of Applied Linguistics*. Wiley-Blackwell. <https://doi.org/10.1002/9781405198431.wbeal1152>