ISSN: 1859-2171
e-ISSN: 2615-9562
TNU Journal of Science and Technology 225(03): 143 - 151

DEVELOPMENT AND VALIDATION OF AN ENGLISH WRITING TEST AT A UNIVERSITY IN VIETNAM

Nguyen Xuan Nghia
School of Foreign Languages, Hanoi University of Science and Technology

ABSTRACT

This study was conducted in an attempt to replace the writing component of an Olympic English test battery at a Vietnamese university. After the test was developed with reference to Bachman and Palmer's test construction model, it was administered to 18 participants at the university. The scripts were then independently marked by two raters, and the scores were used as evidence to determine construct validity and scoring validity of the test and test procedures. The Pearson correlation test was employed to check internal consistency of the test and scoring consistency between the raters. Correlation coefficients R = 0.72 and R = 0.94 suggested that the two test tasks well reflected the writing ability construct defined in the test, and R = 0.43 indicated both an intersection and a discrimination in the content and difficulty level of the test tasks. Inter-rater reliability was recorded at a satisfactory level (R = 0.74), but this value could have been enhanced with more strict marking guidelines applied to problematic scripts.

Key words: *test development; test validation; construct validity; scoring validity; writing ability*

Received: 24/02/2020; Revised: 09/3/2020; Published: 23/3/2020

XÂY DỰNG VÀ XÁC TRỊ ĐỀ THI VIẾT TIẾNG ANH CỦA MỘT TRƯỜNG ĐẠI HỌC TẠI VIỆT NAM

Nguyễn Xuân Nghĩa
Viện Ngoại ngữ - Đại học Bách Khoa Hà Nội

TÓM TẮT

Nghiên cứu này thực hiện nhằm mục đích thiết kế lại đề thi kỹ năng Viết trong bộ đề thi Olympic tiếng Anh tại một trường đại học ở Việt Nam. Đề thi sau khi thiết kế dựa trên mô hình xây dựng đề thi của Bachman và Palmer được tiến hành cho thi trên 18 sinh viên của trường đại học này. Bài viết sau đó được chấm bởi hai giám khảo độc lập; điểm số của các bài viết này được sử dụng để xác định độ giá trị cấu trúc và độ nhất quán đánh giá bài thi. Hệ số tương quan Pearson được sử dụng nhằm kiểm tra độ nhất quán trong nội tại bài thi và nhất quán trong việc đánh giá bài thi giữa hai giám khảo chấm thi. Hệ số tương quan đạt mức R = 0,72 và R = 0,94 cho thấy hai câu hỏi của đề thi đã phản ánh khá tốt khái niệm kỹ năng Viết được xác định trong đề. Đồng thời hệ số tương qua giữa hai câu hỏi đạt giá trị R = 0,43 cho thấy hai câu hỏi vừa có độ nhất quán vừa có độ phân hoá. Sự đồng thuận giữa hai giám khảo cũng đạt mức khá (R = 0,74), tuy nhiên để cải thiện hơn nữa giá trị này cần có quy trình hướng dẫn chặt chẽ hơn đối với các bài viết chưa đạt yêu cầu.

Từ khoá: *xây dựng đề thi; xác trị đề thi; độ giá trị cấu trúc; độ nhất quán đánh giá; năng lực viết*

Ngày nhận bài: 24/02/2020; Ngày hoàn thiện: 09/3/2020; Ngày đăng: 23/3/2020

Email: nghia.nguyenxuan@hust.edu.vn
DOI: https://doi.org/10.34238/tnu-jst.2020.03.2706

1. Introduction

The Olympic English Contest (OEC) at Oxfam University of Hanoi (pseudonym) has been around for nearly two decades now. It serves as a measure of linguistic ability of its freshman and sophomore students, based on which the best scorers are incentivized with prize money, bonus points, and certificates. Its test battery consists of four subtests, corresponding to four English macro skills. While the reading, listening and speaking subtests have marked resemblance to those of the IELTS test, the writing component is rather independent in respect to its content and number of task types, so was purposively chosen for investigation in this study and hereinafter referred to as English Writing Test or EWT. The EWT deals with academic domain of knowledge and contains a single timed task that looks for an extended argumentative essay. The task is structured in a way that a paragraph-length prompt (30-40 words) functions as a lead-in to a guiding question at the end. Its topical area changes every year and is chosen from a repertoire of education, economy, culture, and technology, among others.

Having been operational for such a long time, the EWT has never undergone a formal revision despite a number of issues associated with its validity. First, the fact that it is constituted by a single task does not seem to insure coverage of what is embedded in the real-world setting. In a genuine academic scenario, students are asked to produce not only a discursive text but also varied forms of written communications such as emails or letters. Second, an independent writing task in more recent testing practices is losing momentum to integrated writing in which the composition is accompanied by listening and/or reading requirements. This practice has been partly mirrored in the writing section of the TOEFL test. Furthermore, one of the biggest limitations of the EWT lies perhaps in its scoring method and procedures. Each set of collected scripts is assigned to a random teacher for marking in an impressionistic fashion and is not subject to remarking or second marking. For all of these reasons, I found it worth an attempt to reexamine the current test and redevelop it in a way that its validity is assured prior to use. To this end, the study sought to address two questions:

- To what extent does the new EWT have construct validity?
- To what extent does the new EWT have scoring validity?

2. Literature review

2.1. Test development

Language testing specialists suggest different test construction procedures, depending on purpose of the test (e.g. placement vs. proficiency), type of the test (paper-and-pencil vs. performance), and difficulty level of the test etc. [1],[2],[3,[4]. For example, McNamara works out a four-stage process: understanding the constraints, test design, test specifications, and test trials [4]. Hughes makes a list of ten steps, with making a full and clear statement of the testing problem and training staff such as raters or interviewers on the two ends of the chart [2]. The most full-fledged test development framework perhaps is that by Bachman and Palmer with three stages – design, operationalization, and administration – to which this study was anchored for construction of the new EWT [1].

2.1.1. Test design

This initial stage of the test development cycle targets at a "design statement" in which a host of items, but most importantly purpose of the test, description of the target language use domain and task types, and definition of construct, are displayed. Test purpose can be viewed from three perspectives: types of inferences to be made from test scores, educational decisions made on the basis of test scores, and the intended impact on test users. It is on the first dimension that language tests are based and arranged on a continuum starting with achievement test and ending in proficiency test. The second dimension – educational uses – is foundational to the classification of language

tests into formative testing and summative testing. The last set of test purposes is derived from the range of stakeholders the test may impact, whether it be an individual student or other major parties alike such as teachers, institutions, and society, so corresponds with low-stakes and high-stakes tests [5].

The target language use (TLU) domain is defined 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" [1]. Take the academic module of the IELTS as an example. The TLU domain is determined as an academic university setting, so the writing assignment task, for example, translates in Task 2 of the writing subtest. Through this test task, IELTS test writers are trying to measure the test taker's writing ability, towards capturing the overall picture his or her overall language ability. This ability is an intangible attribute of the test taker and is coined under the term "construct". It is a covert and latent theoretical concept rather than an overt and concrete one [6]. The definition of construct can be attained in light of instructional objectives in a course syllabus or a theoretical account of language ability [7]. In this regard, it is more plausible to fit the construct underlying the EWT in aspects of writing ability. Raimes develops eight features of writing ability, namely content, the writer's process, audience, purpose, word choice, organization, mechanics, and grammar and syntax [8]. Heaton defines one's writing ability through four areas of knowledge: grammatical knowledge, stylistic knowledge, mechanical knowledge, and judgemental knowledge [9].

2.1.2. Test operationalization

The central task in operationalizing a test is to formulate a test specification which functions as a "blueprint" for immediate and future versions of the test to be written [3]. This blueprint provides details about the structure of the test and about each test task, for instance, number and sequence of test tasks/ parts, and definition of construct, time allotment, instructions, scoring method, and rating scales etc. for each task [1]. Of crucial concern to performance tests is scoring method as it has a direct impact on test scores, which are in turn deterministic to validity of the test. There are two commonly used scoring methods – holistic scoring and analytic scoring [7]. Holistic scoring refers to the rater's assigning of a single score to a piece of writing on its overall quality based on his or her general impression [7] [10]. The drawback of this rating method is its inability to make informed decisions about a script as a result of a lack of explicitly stated criteria to be marked against [11] [12]. With analytic scoring, by contrast, the rater judges several facets of the writing rather than giving a single score. A script can be rated on such criteria as organization of ideas, cohesion and coherence, lexical and grammatical resource and mechanics [7]. This is why analytic scoring lends itself better to rater training [7] and reliability enhancement [13].

2.1.3. Test administration

The test administration step in Bachman and Palmer's 1996 framework involves administering the test, collecting feedback, and analyzing test scores. Evidently, it is deficient in test trialing, a step that is salient in other scholars' procedures. This gap has recently been filled in their updated 2010 version with a five-stage process put forth: initial planning, design, operationalization, trialing, and assessment use [14]. Test trialling entails trying out the test materials and procedures on a group of people who, in all respects, resemble the target test population, and where there is subjective marking of speaking and writing, there is a need for training of raters [3]. The feedback, including perceptions on the clarity and comprehensibility of test prompt, level of difficulty of test tasks and so on, is then collected and used to inform adjustments made to the original version of the test. Tryouts can be multiple for minimization of flaws and ambiguities. This is how to make sure the test has been carefully scrutinized before being administered to a larger population.

2.2. Test validation

"Validity refers to the appropriateness of a given test or any of its component parts as a measure of what is purported to measure" [15]. Validity is indexed in three ways: first, the extent to which the test sufficiently represents the content of the target domain, or content validity; second, the extent to which the test taker's scores on a test accurately reflect his or her performance on an external criterion measure, or criterion validity; and third, the extent to which a test measures the construct on which it is based, or construct validity [16]. Validation is the collection and interpretation of empirical data associated with these validity evidences [17]. Content validity evidence can be elicited by interviewing or sending out questionnaires to experts such as teachers, subject specialists, or applied linguistics and obtaining their views about the content of the test being constructed. Criterion validity is performed by correlating the scores on the test being validated and a highly valid test that serves as the external criterion. If this correlation coefficient is high, the test is said to have criterion validity. The achievement of construct validity evidence is grounded on a number of sources, including the internal structure of the test, i.e. the correlation between test tasks/ items, and correlational studies, i.e. correlation of scores of the present test and another test supposed to capture the same construct [17]. As Bachman puts it, the validation of a test cannot be divorced of reliability checks since reliability is a part and parcel of validity [1]. As well, reliability is not constituted by end scores but must be constantly attended to en route. With respect to a writing test, reliability is essentially about consistencies in the ratings of a single rater with scripts of same quality or same scripts and consistencies among different raters [7]. It is this set of statistics that the study looked into, in combination with those on construct validity, in validating the EWT.

3. Methodology

3.1. Participants

The participants were 18 first- and second-year students (N = 18) from School of Foreign Languages, Oxfam University of Hanoi. In April 2019, I visited different classes and familiarized students with my project and my wish to have them as test takers. I did not face any difficulty as all those who volunteered to participate in my study had taken the actual Olympic English Contest a bit earlier that year, so they were even excited to take the new version of the writing subtest. To encourage their commitment to taking the test and doing their best, I promised to offer them stationeries such as pens and highlight pens when the test was done. Of these 18 individuals, there were 13 females and 5 males, with their proficiency levels revolving around upper-intermediate.

3.2. Test development

The development of the new EWT began with the determination of the TLU domain and the content to fit in this criterion. In accordance with Olympic English Contest Development Committee's guidelines, the EWT was made as a proficiency test that simulated academic language tasks from higher education contexts. With a clear test purpose and criterion setting in mind, I referred to literature relative to writing genres and aspects of writing to arrive at a global picture of the writing ability construct. From Hedge's classification of writing into six categories – personal writing, public writing, creative writing, social writing, study writing, and institutional writing [18] – I found study writing, e.g. essays, and institutional writing, e.g. student-lecturer email communications, highly pertinent. To ensure authenticity of the essay task, I opted for a reading-to-writing task with a text-based prompt [19] given that a genuine writing task is frequently planned rather than impromptu and that "university writing is virtually always based on some prior reading" [7]. As with the email chore, it was conformed to framed prompt-driven writing with a circumstance set out for

response or resolution [19]. These two writing tasks, combined with an examination of Raimes's aspects of writing – content, the writer's process, audience, purpose, word choice, organization, mechanics, and grammar and syntax – [8], helped me to decide on the scoring method, which is analytic scoring, what went into the scoring guide, and the test specification as a whole. I capitalized on Jacobs et al.'s rating scheme by virtue of its proven reliability and overlap with aspects of writing ability drawn upon in this study [19]. With considerations of additional task features such as time allotment, instructions and so on, the test specification was in place, providing a blueprint for the test to be written.

3.3. Test trialling

After writing the test on the basis of the test specification, I carried out pretesting procedures, including a pilot test and a main trial, as suggested by Alderson et al. [4]. In order to pilot the test, I involved three native-speaker students and two local students, two males and three females, on Oxfam University of Hanoi campus. My intention was to have them voice their opinions about the comprehensibility and difficulty of the test. After five minutes of reading the test, the students were asked to respond to this list of questions:

- Are there any words you do not understand?
- Do you know what you have to do with this test?
- Are there any particular words, phrases, or instructions that you find confusing and might affect your response?
- Are there any changes you would suggest be made to the test?

All the students thought that it was a "very good" and "easy-to-understand" test. They suggested correcting the phrase "you and other two students" into "you and two other students". Later I used these invaluable comments to modify the wording of the prompt (the final version of the test can be found in Appendix). For further insights, I requested their actual tryout with the test but they all refused because of their lack of time and the length of the test.

In early May, I administered the new EWT to the 18 participants as a main trial. They all gathered on a Sunday morning at a room I had earlier set up and did the test. They wrote their answer on a separate answer sheet and were not allowed to use dictionary and electronic devices. After 70 minutes, I collected the papers and gave away stationeries for their participation. It was this set of scores assigned to these papers that I later analyzed as an initial step of the validation procedures.

3.4. Rater training

Due to time and financial constraints, I was unable to carry out formal training sessions or hired certified raters but involved a friend of mine who was willing to act as a second rater besides myself. He was a teacher at a different institution and shared with me a teaching background and a command of English. We shared an IELTS overall score of 8.0 with a writing sub-score of 7.5 and both had experience teaching writing skills and marking scripts. After the main trial session with the participants, I set up an appointment with the rater. We had talks about the scoring rubric and how to handle problematic scripts such as off-task, unfinished, and under-length scripts (I had read through the scripts once collecting them). We also discussed potentially ambiguous words like "substantive" and "conventions". At the beginning as well as the end of the discussion, I carefully described the test and related issues to him in order to make sure he would mark the test with a clear idea of the context in mind. After that, we independently marked the scripts at our own convenient time for two days, and he returned me the scores on the third.

4. Findings and discussion

4.1. Research question 1: To what extent does the new EWT have construct validity?

The first question this study sought to answer was to what extent the new EWT was valid with regards to its construct, i.e. the construct

of writing ability. First, Pearson product-moment correlation coefficients were computed to check whether the scoring guide was the right choice in this study. The fact that the figures of .98, .95, .96, .95, and .91 for content, organization, vocabulary, language use, and mechanics respectively showed that these components satisfactorily reflected the writing construct under investigation, and the overall scoring guide was reliable.

As pointed out by Bachman, construct validity evidence can be obtained by looking into the internal consistency of the test [17]. Therefore, I examined three relationships – one between Task 1 and the overall test, one between Task 2 and the overall test, and one between the two test tasks – by depending on the Pearson correlation test. The results are shown in Table 1.

The correlation coefficient R = 0.72 suggested that there was a quite strong correlation between students' scores on Task 1 and the overall scores of the test, an indicator of a relatively good representation of the writing construct in this task. The value was significantly higher as with Task 2 – EWT relationship (R = 0.94). This also meant there was up to 80 per cent of the writing construct demonstrated in Task 2. The extent to which Task 1 and Task 2 correlated with one another was noteworthy here. The value R = 0.43, i.e. nearly 20 per cent agreement, revealed that both the tasks reflected the writing construct in each other but also discriminated in level of difficulty.

4.2. Research question 2: To what extent does the new EWT have scoring validity?

The other type of validity the study was interested in was scoring validity which is often referenced as intra-rater reliability and inter-rater reliability [7]. As we, the raters, did not have time to mark a single script twice, only evidence pertaining to inter-rater reliability was unearthed, again by means of the Pearson correlation test. The correlation coefficient was calculated on sets of scores awarded by two raters to the 18 scripts and was determined at R = 0.74. Though this value fell into an acceptable range of 0.7 – 0.9 as suggested by McNamara for inter-rater reliability [3], it was not remarkably high, for it indicated only about 55% of agreement between the raters. So, I attempted to explore where the disagreement might have come from by looking into (1) correlation of scores given by two raters for each scoring criterion, and (2) mean scores given for each scoring criterion.

Table 2 shows that Content was the only area where the raters were in agreement to a significant extent while disagreement of varying degrees occurred with the other four, especially Mechanics. If we look at the means of the overall and component scores awarded by the raters (Table 3), it is fair to say that Rater 2 tended to give higher scores than Rater 1 on every scoring aspect.

Table 1. *Correlations between scores on each test task and the whole test (R)*

	Task 1 – EWT	**Task 2 – EWT**	**Task 1 – Task 2**
Correlation coefficient (R)	0.72	0.94	0.43

Table 2. *Correlation of scores given for each scoring criterion*

Correlation coefficient (R)				
Content	**Organization**	**Vocabulary**	**Language use**	**Mechanics**
0.84	0.62	0.52	0.56	0.34

Table 3. *Mean scores given for each scoring criterion*

	C	**O**	**V**	**L**	**M**	**Overall**
Rater 1	22	14.7	14.3	17.6	4.32	72.5
Rater 2	23.6	15.9	16.2	18.8	4.39	78.8

Table 4. *Scores awarded by two raters to problematic scripts*

Name code	Script problem	Rater 1			Rater 2		
		Task 1	Task 2	Average	Task 1	Task 2	Average
G	Off-task 2	90	58	67.6	82	78	79.2
J	Off-task 1	34	71	59.9	65	80	75.5
L	Incomplete task 2	74	50	57.2	87	73	77.2
O	Under-length task 2	70	48	54.6	60	65	63.5

Another source of disagreement that was worth investigation concerned problematic scripts. During the marking process, I found three types of problems with the students' writings: off task, incomplete, and under-length, to name them. This is congruent with Weigle's caveat of potential issues that might affect scoring validity [7]. The scores assigned to these pieces of writing are presented in Table 4. Table 4 uncovers the reality that there were discrepancies of differential yet large degrees in scores awarded to problematic writings. For example, while Rater 1 assigned only 58 points for Task 2 written by participant G for his misinterpretation of the instructions and/or questions, Rater 2 gave up to 78. This was the same case as with Task 1 misconstrued by participant J. This leads to the question of to what extent the raters understood what it meant by off-task by themselves and to what extent they understood each other during the rater discussion session. With respect to the other problematic scripts, scores were also awarded differently, demanding the raters to have communicated more openly and effectively prior to the marking.

5. Conclusion

This study aimed to develop and validate the writing subtest of the Olympic English Contest test battery at Oxfam University of Hanoi. Though the test was neither developed from scratch nor validated in light of a comprehensive validation framework, it underwent major procedures of a test construction cycle and was validated with empirical data from a trial test with quite a few participants. The study came to the conclusion that the reconstructed test achieved a high level of construct validity, and the raters were in agreement in scoring. Having said that, the findings suggested that rater training have been implemented in a more formal and strict fashion to avoid misinterpretations of any details in the scoring guide and writing issues such as off-task, incomplete, and under-length scripts. For example, there should have been common grounds on how many points a problematic script could get at a maximum. I am aware that the present study was yet to produce a perfect test as the population on whom the test was tried out was not tens or hundreds, and other aspects of validity demanded for investigation in more depth and breadth, this is a task that will be performed if the test is put to use in near future.

REFERENCES

[1]. L. F. Bachman and A. S. Palmer, *Language testing in practice: Designing and developing useful language tests*. Oxford: Oxford University Press, 1996.

[2]. A. Hughes, *Testing for language teachers.* Cambridge: Cambridge University Press, 1989.

[3]. T. McNamara, *Language testing*. Oxford: Oxford University Press, 2000.

[4]. J. C. Alderson, C. Clapham and D. Wall, *Language test construction and evaluation*. Cambridge: Cambridge University Press, 1995.

[5]. S. Stoynoff and C. A. Chapelle, *ESOL Tests and Testing: A Resource for Teachers and Administrators*. Alexandria, VA: TESOL Publications, 2005.

[6]. L. J. Cronbach and P. E. Meehl, "Construct validity in psychological tests," *Psychological Bulletin*, vol. 52, no. 4, pp. 281-302, 1995.

[7]. S. C. Weigle, *Assessing writing*. Cambridge: Cambridge University Press, 2002.

[8]. A. Raimes, *Techniques in teaching writing*. New York: Oxford University Press, 1983.

[9]. J. B. Heaton, *Writing English language tests*. New York: Longman Inc., 1994.

[10]. A. Davies, A. Brown, C. Elder, K. Hill, T. Lumley, and T. McNamara, *Dictionary of language testing*. Cambridge: Cambridge University Press, 1999.

[11]. P. Elbow, "Writing assessment in the 21st century: A Utopian view," in *Composition in the 21st Century: Crisis and Change*, L. Z. Bloom, D. A. Dailer, and E. M. White, Eds. Carbondale: Southern Illinois University Press, 1996, pp. 83-100.

[12]. B. Hout, "The literature of direct writing assessment," *Major concerns and prevailing trends. Review of Educational Research*, vol. 60, no. 2, pp. 237-263, 1990.

[13]. L. Hamp-Lyons, "Pre-text: Task-related influences on the writer," in *Assessing second language writing in academic contexts*, L. Hamp-Lyons, Ed. Norwood, NJ: Ablex, 1991, pp. 69-87.

[14]. L. F. Bachman and A. S. Palmer, *Language assessment in practice*. Oxford: Oxford University Press, 2010.

[15]. G. Henning, *A guide to language testing*. Cambridge: Newbury House, 1987.

[16]. W. J. Popham, *Classroom assessment: What teachers need to know*. Boston: Allyn and Bacon, 1995.

[17]. L. F. Bachman, *Fundamental considerations in language testing*. Oxford: Oxford University Press, 1990.

[18]. T. Hedge, *Writing*. Oxford: Oxford University Press, 2005.

[19]. B. Kroll and J. Reid, "Guidelines for designing writing prompts: Clarifications, caveats, and cautions," *Journal of Second Language Writing*, vol. 3, no. 3, pp. 231-255, 1994.

APPENDIX: THE NEW EWT

Task 1

You and two other students as a team are preparing a presentation on the topic: ***"A traditional festival in your country that you know, have attended or would like to discover about"***. Before presenting, you need to obtain your lecturer's approval of the suitability of your topic and ideas. As a team leader, you are going to write an email to your lecturer, briefly describing the structure of your presentation.

As you write your email,

- start with "Dear Assoc. Prof. Jamie," and end with "Sam" instead of your real name.
- include at least three points that will be used for your presentation.
- you can use bullet points but must write in complete sentences.
- you should write at least 150 words.
- you should spend about 20 minutes on the task.

Your writing will be assessed on content, organization, vocabulary, language use, and mechanics. This task counts for **30%** of the total mark.

Task 2: *First, read the following passage:*

The Homework Debate

Every school day brings something new, but there is one status quo most parents expect: homework. The old adage that practice makes perfect seems to make sense when it comes to schoolwork. But, while hunkering down after dinner among books and worksheets might seem like a natural part of childhood, there's more research now than ever suggesting that it shouldn't be so.

Many in the education field today are looking for evidence to support the case for homework, but are coming up empty-handed. "Homework is all pain and no gain," says author Alfie Kohn. In his book *The Homework Myth*, Kohn points out that no study has ever found a correlation between homework and academic achievement in elementary school, and there is little reason to believe that homework is necessary in high school. In fact, it may even diminish interest in learning, says Kohn.

If you've ever had a late-night argument with your child about completing homework, you probably know first-hand that homework can be a strain on families. In an effort to reduce that stress, a growing number of schools are banning homework.

Mary Jane Cera is the academic administrator for the Kino School, a private, nonprofit K-12* school in Tucson, Arizona, which maintains a no-homework policy across all grades. The purpose of the policy is to

make sure learning remains a joy for students, not a second shift of work that impedes social time and creative activity. Cera says that when new students are told there will be no homework assignments, they breathe a sigh of relief.

Many proponents of homework argue that life is filled with things we don't like to do, and that homework teaches self-discipline, time management and other nonacademic life skills. Kohn challenges this popular notion: If kids have no choice in the matter of homework, they're not really exercising judgment, and are instead losing their sense of autonomy. (Johanna, 2013)

** "K12, a term used in education in the US, Canada, and some other countries, is a short form for the publicly-supported school grades prior to college. These grades are kindergarten (K) and grade 1-12." (whatis.techtarget.com)*

Now, write an essay answering the following questions:

1. ***What is author's point of view?***
2. ***To what extent do you agree or disagree with this point of view?***

As you write your essay,

- follow an essay structure (introduction, body, and conclusion).
- write in complete sentences.
- explicitly address both the questions.
- you can use the ideas from the passage but must rephrase and develop them.
- balance the author's viewpoint and your own.
- provide relevant examples and evidences to support your points.
- you should write at least 250 words.
- you should spend about 50 minutes on this task.

Your writing will be assessed on content, organization, vocabulary, language use, and mechanics. This task counts for **70%** of the total mark.