

How Validity is Ensured in Our Language Test: a Case Study.

Mohammed Humayun Kabir*

Abstract: *In this paper we will first attempt to familiarize the paradigm shift in language teaching in Bangladesh. Then, the umbrella term validity is discussed with references to the literature. Subsequently, three different data instruments have been analysed in order to investigate how far validity is maintained in testing Reading and Writing taking HSC (Higher Secondary Certificate) English testing as sample. The study found that validity at HSC English Testing is not maintained as item setting, marking, etc. are problematic. Finally, this study recommends some pragmatic measures like ensuring question setter's qualifications and orientation, bringing some changes in the item setting, providing rater's training, etc. to achieve the desired goal.*

Introduction:

In Bangladesh our Ministry of Education has brought a ground-breaking change in English language teaching and testing system by shifting from previously followed grammar –translation-literature based syllabus to CLT (Communicative Language Teaching) approach. CLT was introduced in Secondary and Higher Secondary Certificate (HSC) curriculum in 2000-2001. In order to make language teaching effective we needed experimentation and evaluation of the curriculum. Candlin and Widdowson in Nunan (1988a :ix-x) write '*If language teaching is to be a genuinely professional enterprise, it requires continual experimentation and evaluation*'

Though CLT insists on communicative performance, we still notice that our students are very busy with preparing answers to the probable questions and memorizing them in order to ensure pass mark (33%) or better Grades in the final exams. This tendency is ever increasing as

* Assistant Professor, ELL, IIUC

our test items are too **‘predictable’**. ‘Our examination system (setting of questions, marking papers , learning outcome, etc) needs to be modified ...’ opines (**Rozario**; 2005). Moreover, it is noticed that the contents of the test are also not judiciously chosen. All these aspects are obviously a validity concern. A researcher, (**Islam**; 2003:15) finds ‘they(teachers) do not have either pre-service or in-service training provision in government policy’. It is the lack of teacher’s orientation more particularly lack of teachers’ training that is largely hampering the English teaching and testing .

I was deeply interested to investigate the validity in our language testing especially at HSC level, as I was a teacher cum examiner without having any orientation in language testing. Therefore my core research question in this study is –How far is validity maintained in testing Reading and Writing skills in HSC English testing?

There now follows a short out -line of marks allocation for HSC English and types of testing:

1.1 HSC English marks distribution:

Marks allocation for different parts in Eng 1st paper

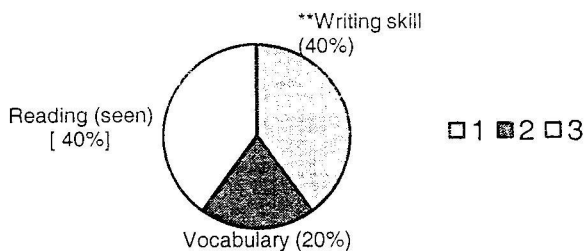


Figure:1.1, [**writing components : Re-arranging, substitution table, paragraph writing.]

Marks allocation for different parts in Eng 2nd paper

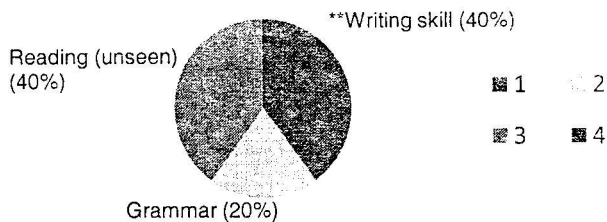


Figure:1.2, [**writing components : Paragraph writing, letter writing / writing creatively from experience, completing story / continuing passage]

1.2 Scoring:

At HSC students are awarded a Grade (A+, A, A-, B, C, D, F) making a total of that s/he scores in Reading test, Vocabulary test and Writing test combined. So we see that global rating scale is followed here.

1.3 Purpose of language test:

After teaching, the question of testing comes very naturally. Rea (1978) (cited in Carroll) suggests

‘communicative language tests for personal and professional purposes imply a set of testing ‘tasks’ which reflect and stimulate actual linguistic demands within specified domains’.

(Carroll :1980:5-6) finds

‘the interactive relationship between the language programme and the testing system, and their mutual derivation from an analysis of the learner’s communicative needs’.

We will look for how far HSC testing reflects these characteristics.

It is true that testing language ability is very difficult. Very often students’ true language abilities are not reflected in their tests’ results. (Hughes; 2003) identifies two main sources of inaccuracy in language ability testing, i.e. ‘test content and test techniques and lack of reliability’. So contents of the tests are vitally important factor in deciding the language abilities. This is known as ‘**content validity**’ .

2. Literature review:

In this we will discuss the umbrella term *validity* , the main issue of concern in our study.

2.1 Validity:

Validity of a language test refers to what a test actually tests in relation to what it is supposed to test. It is ‘the most important question of all in language testing’ (Alderson et al, ibid:170). (Henning; 1987:89) states validity ‘refers to the appropriateness of a given test or any of its component parts as a measure of what it is purported to measure’, cited in (Alderson et al,1995). The concern for validity has been a very familiar and popular issue for a quite long time. (Kelly,1927:14) stated ‘the problem of validity is that of whether a test really measures what it purports to measure’, cited in (Weir, 2005:12). (Lado, 1961:321) similarly raised a question and then finds the answer in the

following manner, 'does a test measure what it is supposed to measure? If it does, it is valid', cited in (Weir, 2005:12). Almost a similar stand point is noticed when Cronbach (1971:463) opined 'every time an educator asks 'but what does the instrument really measure?' he is calling for information on construct validity', cited in Weir (2005:12). Hughes (2003:26) has confined the whole term validity under 'construct validity'. He states 'in recent years the term construct validity has been increasingly used to refer to the general, overarching notion of validity'.

2.2 Content validity:

'Content validity is the representativeness or sampling adequacy of the content –the substance, the matter, the topics- of a measuring instrument' Kerlinger (1973:458), cited in Alderson (1995:173). Anastasi (1988:131-32) and Hughes (2003) also have fostered almost similar sort of idea regarding content validity. But Anastasi (1988) insists that 'all major aspects are covered by the test items, and in the correct proportion' when he refers to it (content validity), cited in Weir (2005). If we find in a test the representative sample of the language skills, structures, etc. with which it is concerned then we can claim that this test has **content validity**. If it is a communicative language test and if we are testing reading skills or writing skills, the test must have the salient representative samples of which communicative language test is concerned with. Hughes (2003:27) states 'a comparison of test specification (see section 2.4) and test content is the basis for judgments as to content validity'. Of course this judgement must come from '**expert**': 'people whose judgment one is prepared to trust, even if it disagrees with one's own', Alderson et al, (1995:173). It is accepted that a higher content validity can ensure a greater construct validity.

According to Hughes (2003) the second form of evidence of a test's construct validity is 'criterion –related validity', a predominantly quantitative concept which is 'concerned with the extent to which test scores co-relate with a suitable external criterion of performance (see Anastasi (1988:145, Messick (1989:16) with established properties' cited in Weir (2005:35). Hughes (2003) necessitates two more types of validity –validity in scoring and face validity.

This research will investigate the extent to which the HSC exams are **valid in terms of content**.

2.3 Validity in scoring:

If we simply consider the items (content validity) and overlook the way in which the responses are rated or scored, then the validity of test will never be established. 'It is no use having excellent items if they are scored invalidly', **Hughes** (2003:32). For example in reading skill test a rater should not take into account spelling and grammar, if he takes these features into account while marking, then it is not a valid measurement. **Weir** (2005:23-24) states 'scoring validity seems to be the most suitable and if a test lacks this its validity is seriously threatened'. At HSC the validity of scoring is a matter of great concern specially when there is subjective marking . So it is an important issue in this research.

2.4 Test specification:

A test's specifications are official documents from the responsible authority which narrate what the test tests and how it tests . 'They are also essential in the establishment of the test's **construct validity**' (**Alderson** et al, 1995:9). In this statement one can find the rationale, contents, and marking schemes of a test. It is equally helpful and essential for test writers, test takers, teachers, and markers as well. 'Test specifications are also needed by those responsible for or interested in establishing the **test's validity**', (ibid:10).

3. Instruments:

In order to find the answer of my research question I will use the following instruments:

- a) One questionnaire addressed to the selected teachers & examiners (45) of higher secondary schools and another questionnaire addressed to the Chairman / Controller of examination of Chittagong Board of Education (CBE).
- b) Official Instructions to HSC English examiners.
- c) Previous test papers(Board questions) .

In addition to these above written documents, I used some more official papers and reports to support my findings in my research with official information.

Data analysis:

4.1 Reply to the 1st Questionnaire:

Among 45 questionnaires I got back 19. In those data I found that the teachers' / examiners' experience varied from 1 year to 18 years. I got reply from the teachers belonging to the colleges situated in the city

However, the questionnaire data reveals that almost 63% teachers use T.G. May be they are using it (partially) for different purposes like: making lesson plan, setting questions, preparing answers for privately taught students, etc.

At the end of the T.G. there is valuable discussion on testing under the following titles : 'Testing in the new syllabus' and 'Notes for Teachers, Question setters and Markers' as well . In order to make the **test valid** and the test's result reliable these instructions demand a very thorough and attentive reading and internalisation.

The data gives us some hints that items in the question papers are very much predictable and hence the teachers are under pressure to provide suggestions and answers of the probable questions.

In the questionnaire data (Q.14) it is also depicted that marking varies significantly from markers to markers. 94.73% participants acknowledged it.

Q.18 asks about whether the examiners support the present testing (marking) system, 57.89% respondents supported and 42.11% expressed a different view.

Responses to open-ended questions:

There are 7 open – ended questions in the questionnaire for the teachers.

When asked 'what qualifications are required to be a question setter' (Q.17), 5 participants did not reply, perhaps they did not have any information. 4 participants directly and honestly acknowledged that they knew nothing about it. 3 participants provided their own opinion:

T5: 'To be honest and at home in that particular subject'.

T7: 'comprehensible and give chance the students to show a little bit creativity'

But only 2 participants seem to provide authentic information:

T4: 'minimum status of a teacher is to be an Assistant Professor'.

T11: 'mid senior to senior position'

Q.4 asks about the specific criteria to be an examiner. 5 participants could not provide any information; they simply wrote '**I do not know**'. 1 participant did not respond. 4 participants wrote their own opinion (not the education board provided criteria), though all of them are marking scripts regularly.

T4 : ‘Teacher of a recognized institution and must have experience and sincerity’. This reply seems to me self-contradictory as the information he provided in the questionnaire data reveals that he has been marking HSC scripts for 10 years and his teaching experience is 10 years as well. Moreover, in my study there are respondents from various institutions which even cover some very little known institutions. However, T8 provides some shocking information when she writes that apart from long experience ‘good relationship with (Education) board authority’ is needed to be an examiner. If it is so , we presume that the item setting will be very much problematic. As a result the validity of the test is supposed to be at stake.

Inquiry related to marking reading skill (Q.8):

When the markers were asked what qualities they look for while marking reading comprehension – 6 participants wrote that they look for capacity for understanding, 2 participants look for vocabulary, while some others provided puzzling responses:

Inquiry related to marking writing skill(Q.9):

9 respondents are looking for grammar skill e.g. –

T4 : ‘grammatical and language PERFECTION’

While the following look for:

T11: ‘clarity of thought, presentation, gradual development of ideas, coherence, originality’

T2 : ‘ability of writing answer in simple English’.

T5: ‘appropriate answer in simple and own English’.

In the above analysis it is almost established that HSC testing is in grave situation. T.G. which is a very useful tool is not properly used by the teachers. The items of the test are very much predictable. So teachers are providing suggestive question – answers or they are under pressure to supply them. They are still looking for grammatical accuracy in stead of communicative competence. Here we can assume the lack of **validity** in the test (see section 2.1, 2.2 &2.3).

4.2 Reply to the 2nd questionnaire:

Q.5.asks: Do you think marking (significantly) varies from examiner to examiner? If yes, please write why it happens.

He replies : **‘I think it does not occur usually’.**

In the response of Q.7, (Do you think double marking of each script is possible at HSC? If not, please mention why not?)

He writes: '**No. Shortage of time, manpower. Financial constraint. Since marking does not does not vary significantly from examiner to examiner.'**

It is hard to accept the responses because both teachers / markers and daily newspapers' report do not support it. We assume, as he is holding a very responsible position he cannot admit all the anomalies and irregularities exist.

The Education Ministry so far has not been successful establishing 100% free and fair test administration across the country. So we can perceive that HSC testing is seriously lacking in **reliability and hence validity** as well. Since the number of respondents is very small, the results can only be indicative, rather than definitive.

4.3 An official marking Instructions to HSC English examiners:

As usual every year after the ending of the HSC examination every marker is provided with a written document i.e. '**Instruction to Examiners**' by the concerned education board authority. In this study I have taken the 'Instruction to Examiners 2006 provided by CBE as a sample instrument to analyses.

Instruction-h, states 'not more than 50% marks should be awarded for memorized answers. Unintelligent, crammed and bad reproduction should be penalized'. Here the problem lies, I assume, not with the writing of answers, but with the setting of questions. It is claimed (in the syllabus) that the testing should be in line with the Communicative Language Testing. If it is so, why does the question of memorizing or cramming arise? Why should the students be penalised for the inability of the item writers? So there remains a serious doubt in the **validity of testing**.

Further discussion will be made below .

4.4 Previous test papers .

Judicious item setting is a vital issue in language testing especially with regard to content and construct validity. If there are methodological problems in terms of item selection , if items are not selected in line with the ideal of Communicative Language Testing , if the items become predictable , if the content validity is not maintained , then both the validity and reliability of testing will be in doubt or even the test will lose its acceptability among the test users. After analysing the HSC Board questions of Chittagong and Dhaka boards

of 2004, 2005, 2006, of English 1st Paper we have found the following items in the reading and writing skill testing

Analysis of the English 1st Paper Question 2006 of Chittagong Education Board:

4.4.1 Reading skill testing:

Q.1 to Q.4 are to be answered after reading 1st passage [cited from Unit-23, Lesson-(2B) of the textbook.]

Q.1 (of 5 marks) consists of five MCQ type items. 'Multiple – choice questioning can be used effectively to train a person's ability to think It is possible to set distracters so close that the pupil has to examine each alternative very carefully indeed before he can decide on the best answer....,' **Munby** (1968:xii-xiii). So the value and **validity** of MCQ testing relies in setting distracters that demand great care from the test takers. In the study it seems that all of the five items are problematic, as they do not test what they suppose to test. We are taking 1 (c) & 1 (d) as samples for discussion.

Q.1 (c): **'The cause of malnutrition in Bangladesh is poverty / negligence / indifference of the government.'** In the given passage the answer is almost directly given. A sentence goes like this: 'Poverty has been an ever – present spectre. As a result, nearly seventy percent of the population suffer from various degrees of malnutrition'. The first sentence starts with 'poverty' and the following one starts with 'as a result'. A student having minimum understanding capacity can effortlessly predict that the correct alternative answer will be 'poverty'.

Q.1 (d): **'Bangladesh was under Pakistani rule for 30/24/70 years'**. It seems that the item writer wants to judge simple arithmetical capacity of the students, not the linguistic ability or competence. The very second sentence of the passage is: 'prior to that, it experienced British colonial rule until 1947, followed by an oppressive existence as East Pakistan from 1947 to 1971'.

It is, therefore, evident that almost the whole Q1(a,b,c,d,e) is seriously lacking in **validity** in terms of item setting in line with Communicative Language Testing. **Alderson** (2000:212), rightly says 'to write plausible but incorrect options that will attract the weaker reader but not the better reader is far from easy'. **Hughes** (2003:76-78) appreciates MCQ testing because 'scoring can be perfectly reliable'. However, he is very concerned about the demerits of this testing. He opines MCQ testing 'tests only recognition knowledge', so it 'may give a quite inaccurate picture' of the candidates' ability.

Q.2 (a, b, c, d & e) is also an objective test with dichotomous items (true/false). Here students are asked to give correct information for the false statements. 2 (a), (d) & (e) appear problematic. So those are discussed below.

Q.2 (a): **‘Our political social conditions have positive influences on our society’**. Any student will say it is false and might effortlessly get the correct information directly from the given passage. As it is written ‘turbulent politics and instability and rise of terrorism and crime are adversely affecting society’.

Q.2 (d): **‘our huge population helps to remove poverty’**, does not require to infer correct information from the given passage at all as it is a question related with commonsense. Moreover, it is the answer of Q1(c) where the true information lies.

Q.2 (e): **‘Our Victory Day has got the recognition of UNESCO’**. Any Bangladeshi student at least at HSC level will write that it is false for two reasons: firstly, they know the correct information from their general knowledge, secondly, in the given passage it is written, ‘Our Language day has been declared International Mother Language Day by UNESCO’.

Like Q.1(a,b,c,d,e), Q.2(a,b,c,d,e) is also lacking in **content validity** as both of the items are not properly testing the reading ability of the students.

Q.3 (a,b,c,d,e): ‘Fill in the gaps with the correct form of the words in brackets. Add appropriate prepositions if necessary.’ It demands both language ability and linguistic competence from the student. Hence it is far from Reading Skill testing. From the given ‘key’ (appendix-), it appears that students’ grammatical ability is the main issue of concern behind setting these items. Alderson (2000:98) also shares my view, when he raises the question ‘If grammar and reading tests overlap to a large extent, why bother testing ‘reading’ (although other might say ‘grammar’)? The validity of a reading test might well be undermined if it were shown to measure little more than grammar, if one believes that reading involves much more than the use of linguistic knowledge’.

Q.5 to Q.8 are to be answered after reading 2nd passage [cited from Unit-21, Lesson-4(B) of the textbook.] Q5 (five short answer questions), Q7(a summary writing) & Q.8(making a flow chart), are subjective and Q6(fill in the gaps), is objective in nature. So here again there are reliability concerns during marking (see-4.6). We can predict serious marking problem in Q8(making flow chart), as it is very controversial in nature and it has no absolutely fixed answer. Marking

Q5(short answer questions), will also be problematic. Now we will see what the marking instructions say : Marking instructions to examiners:

No. of the questions	Marking instructions
Q.No. 5	Full marks may be given for satisfactory and accurate answers.
Q.No.6	'Key' is given.
Q.No.7	Summary written in examinees own language / words should given full credit.
Q.No.8	Correct information written in short phrases within the boxes should be assigned full marks.

Table 1.2

It is acknowledged that 'the objectivity of scoring depends upon the completeness of the answer key.....', (Alderson, 2000:227). Marking the answers of the Q.5, Q.7 & Q.8 is beset with the problems as '**Instructions to Examiners**' do not give any specific instructions.

4.4.2 Writing test:

Writing skill is tested in the answers of Q.11, Q.12 & Q.13. In Q.11 (**substitution table**), students have to make six meaningful sentences from this given table. If the sentences are correctly formed but sequence is not maintained full marks might be awarded. This task may be different in nature, e.g. students might be asked to form questions from answer supplied or they might be asked to complete a dialogue where some questions and some answers are supplied and some are missing. Students are to provide the missing bits. Though there is a provision for different patterns of testing writing skills, we have never seen the latter types of tasks set in exam question.

In Q.12 (**re-arranging**), students are to re-order or re- arrange 14 sentences where they have to maintain sequence. If the sequence is broken with any sentence or sentences marks are deducted for that sentence or sentences.

Q.13 (paragraph writing), is answering questions in a paragraph where a topic related to a student's own experience is given which contains four to six questions.

There is no marking guideline to mark Q13 in the marking instructions. It is felt that examiners need some marking guidelines to assess this answer.

Here it is noticed that the test operations are inappropriate. If we carefully observe the tasks, we will find that other than Q.13, students need not produce any creative task. Students are just tailoring the bits or re-arranging them in Q11 and Q12. It is very surprising that the only real writing task (Q13) has a strict word limitation and no marking guide. Students are to produce the answer within 100 words. In fact, we cannot deny that the length of task is an important criterion to judge the student's writing ability, but 2 of the 3 items of writing are not communicative. Hence it is a **validity worry**.

The study also discovers that Q.11 & Q.12 have directly been quoted from the questions of Comilla Board 2005 and Jessore Board 2005 respectively and Q.13, is set from Jessore Board 2004. From this evidence it is obvious that the HSC **test items are becoming very much predictable** and the **validity** of testing is in doubt.

It is categorically written in T.G , 'although comprehension passages are taken from the textbook , the questions (objective and more free) are not'. But Q.5 (a), (c), and (e) are copied from the textbook, Q.5 (d) is also set with a minor change and Q.7, is also quoted from the textbook (pp-272). Since the items are set either from the textbook or from previous question papers students will be allured to memorize the answers and the objective of introducing CLT is bound to be jeopardised. As a result we have reasons to doubt the **validity** of the test.

4.5 To turn now to the HSC Board questions of Chittagong and Dhaka Boards of 2004, 2005, 2006, of **English 2ND Paper** we have found the following items in reading and writing skill testing :

4.5.1 Reading test:

It is almost similar to the English 1st Paper in terms of item selection .In the 'key' for Q.5, it says 'answer to each question should be in one sentence, if written in more than one sentence , it may be given 50%marks'.What does 'one sentence' mean? How will the markers follow it? In '**Areopagitica**' Milton writes sentences which are longer than a normal paragraph. Students having good command over English language may produce longer sentences by using connectors. What will be done with those answers? Another group of students who are unable to write longer sentences will produce very short but correct sentence. So we assume that validity in scoring will be threatened .The weakness of content **validity** that we noticed in the data analysis of reading test of English 1st Paper is also prevalent here.

4.5.2 Writing test:

All education boards have uniformity in testing writing skill. They set the following items: paragraph writing, continuing a passage/completing a story, letter writing, writing creatively from experience. From the above analysis it is evident that HSC Reading and Writing tests lack variety as the items selection for both papers are very common and they are very predictable. We also find that the items of these questions **lack face validity** as most of them do not appear to be real life communicative tasks. As the item writers are only setting those items which are easy to set the **content validity** is also a matter of great concern.

In the above data analysis it is confirmed that at HSC:

1. Question setters are not properly and adequately trained.
2. Most items do not permit objective scoring.
3. No detailed scoring guideline is provided.
4. Items in the question are not set in line with CLT.

So we also find that the **validity** of testing is an issue of great concern as the item selection is seriously flawed.

5. Recommendations:

Here we present the recommendations in the light of the findings of the study.

5.1 Variety in item selection:

After the data analysis (section- 4.1, 4.4.2), we came to a decision that HSC test items are very predictable and hence students increasingly depend on suggestive questions and memorizing the answers of those questions. We advocate ' test developers deliberately vary the methods used so that no one method predominates and becomes predictable', **Alderson** et al's, (1995:46). In the reading skill testing the following techniques might be added to the existing items.

- i) Multiple – matching
- ii) Dichotomous items with third category such as : not given/ the text does not say
- iii) Editing tests
- iv) The gapped summary tests. (Adopted from (**Alderson**, 2000 .cf. Chapter 7) .

In writing skill testing we recommend inclusion of the following items with the existing items:

- i) A series of pictures can be used to elicit a narrative
- ii) Transfer of information from graphic form to continuous prose (Adopted from **Hughes** (2003 .cf. Chapter 9).

While selecting those items it should be remembered that 'some prompts limit the amount of reading required by using a graph or a picture or a series of pictures. In this case it is essential that the graph is easy to understand and that the pictures are clear', Alderson et al, (1995:59).

5.2 Question setter's training:

Question setter's training is urgently needed as most of the so-called experienced question setters of different famous colleges are not well - equipped with CLT. They need urgent orientation in order to construct valid tests.

5.3 Limitations of the research:

It is perceived from the applied nature of this study that examination hall observations , observation of the markers' orientation program, interview of examiners, Head Examiners, question setters could be very useful instruments along with questionnaire (cf. McDonough and McDonough:1997; Nunan:1992; Hopkins:2002).

The Chairman of the CBE was contacted several times, but he could not manage time to respond to the questionnaire as he was tremendously busy with the ongoing HSC 2007 exam. However, the Controller of examinations generously cooperated and responded in time. The findings of the research could have been much thought provoking and insightful if both of them had participated. As HSC is a very large-scale exam, the findings would have been more realistic and authentic if I could have covered all of the Education Boards in the country instead of one.

5.4 Conclusion:

Despite the limitations, it may well be concluded that the findings of the research provide valuable information about the reliability and **validity** at HSC Reading and Writing skill tests. Indeed, the relationship between rater training, rating scale and rater reliability is well understood. To improve the current conditions apart from raters , item writers' watchfulness is also strongly felt. However, if possible future study should include more areas of investigation specially on formulating an exclusive rating scale for assessing reading and writing and effectiveness of rater training in Bangladesh.

Bibliography:

Alderson, J.C.; Clapham, C & Wall, D.(1995): Language Test Construction and Evaluation, CUP

Alderson, J.C (2000): Assessing Reading, CUP

Allison, D. (1999): Language Testing & Evaluation. Singapore University Press.

- Anastasi, A. (1988): *Psychological Testing* (6th edition). New York: Macmillan.
- Carroll, B.J. (1980): *Testing Communicative Performance*. London: Pergamon
- Cronbach, L. J. (1971) : *Test Validation*. In R.L.Thorndike(ed.), *Educational Measurement* (2nd edition). Washington D.C.: American Council on Education.
- Henning, G. (1987) : *A Guide to Language Testing*, Cambridge, Mass. : Newbury House.
- Hughes, A. (1988): *Achievement and Proficiency :The Missing Link*. In Hughes, A.(ed.), *Testing for University Study, ELT Documents 127*. London: Modern English Publications.
- Hughes, A. (2003): *Testing for Language Teachers*. Cambridge University Press.
- Islam, Z. (2003): '*Bridging the Gap: Curricular Innovation and Teacher Preparation Perspective in Bangladesh*.' Dissertation done in the University of Essex.
- Kelly, T.L. (1927) : *Interpretation of Educational Measurements*, New York: New World Book Company.
- Kerlinger, F.N. (1973): *Foundations Of Behavioural Research*. New York: Holt, Rinehart and Winston.
- Khan, R.A. (2005) : Seminar Paper ;3rd International Conference on: 'The Effective Teaching of English in Bangladesh: Policy, Pedagogy and Practices'. Organized by Department of English, Stamford University, Dhaka, Bangladesh. Fall 2005.
- Lado, R. (1961) : *Language Testing*, London : Longman.
- McDonough, J. & McDonough, S. (1997): *Research Methods for English Language Teachers*, Edward Arnold Ltd., London.
- McNamara, T. (2000) : *Language Testing*. Oxford University Press.
- Messick, S. (1989) : Validity. In R.Linn (ed.), *Educational Measurement*. New York : Macmillan pp-13-103
- Nunan, D. (1988 a) : *Syllabus Design* OUP.
- Rea, P. M. (1978): '*Assessing Language as communication*', MALS Journal, University of Birmingham.
- Rozario, B .J. (2005): '*Teaching of English in Bangladesh in Secondary and Higher Secondary Levels*' Seminar Paper ;3rd International Conference on: 'The Effective Teaching of English in Bangladesh: Policy ,Pedagogy and Practices'. Organized by Department of English, Stamford University, Dhaka, Bangladesh. Fall 2005.
- Weigle, S.C. (2002): *Assessing Writing*. Cambridge University Press.
- Weir C. J. (2005): *Language Testing and Validation* . Palgrave Macmillan.
- Textbook:**
- English For Today, For Classes 11-12, NCTB, July-2001; Dhaka, Bangladesh.
- Teacher's Guide for Higher Secondary Certificate English Teacher.

heart to the remotest places (Ranga Mati, Khagra Chori ,West Bansh Khali, etc) under Chittagong education board. I introduce categorization phase of the questionnaire data on the basis of **closed – ended questions** below. In the data analysis Teacher-participant will be mentioned as T1,T2...etc. for the sake of their anonymity in the below .

Title: Teachers’ responses to questionnaire, a summary:

Questions	Yes	%	No	%
1. Teachers are following Teacher’s Guide(T.G)	12	63.15%	7	36.85%
2. Teachers are providing suggestive questions and notes.	14	73.68%	5	26.32%
3. Questions follow syllabus guide-lines.	15	78.95%	4	21.05%
4. Equal standard of test administration.	9	47.36%	10	52.64%
5. Marking varies examiner to examiner	18	94.73%	1	5.27%
6. Following any marking guide	11	57.89%	7	36.85%
7. Whether support present testing(marking) system	11	57.89%	8	42.11%

Table –1.1

From the above table we find that 63.15% teachers are following T.G. While 36.85% teachers honestly confessed that they do not follow T.G. Those who are using it have written that they are using for teaching purpose except T4.He writes **‘sometimes I need to follow T.G. Because it gives the directions how to prepare students for examination’**.

Below, I will discuss the usefulness of T.G in details. Undoubtedly, it helps teachers in teaching but none can deny that it is a useful book for examiner/question setters in the context of Bangladesh. So it can contribute significantly to validate the test. To the reply of the questionnaire-1 teachers acknowledge that they provide suggestions and notes of the probable questions.

As training facilities are insufficient, this T.G. can be a very valuable tool for the practicing teachers. **Khan (2005)**, an ELTIP teacher trainer and researcher points out that,

‘teachers are over burdened with heavy work loads with little time to spare for lesson planning, class preparation or correction of written work. Access to teacher’s guides is nil.....’.