

Developing An English Language Textbook Evaluation Checklist

Jayakaran Mukundan, Universiti Putra Malaysia, Malaysia
Reza Hajimohammadi, Universiti Putra Malaysia, Malaysia
Vahid Nimehchisalem, Universiti Putra Malaysia, Malaysia

ABSTRACT

The paper describes the considerations that were taken into account in the development of a tentative English language textbook evaluation checklist. A brief review of the related literature precedes the crucial issues that should be considered in developing checklists. In the light of the previous evaluation checklists the developers created a list of the evaluative criteria on which the construct of the checklist could be established. The developers considered matters of validity, reliability and practicality in the process of its design; however, further research is in process to refine the checklist. Such an instrument could be used by curriculum designers, material developers and evaluators, as well as English language teachers.

Keywords: English language material evaluation; textbook evaluation checklist

INTRODUCTION

According to Sheldon (1988), we need to evaluate textbooks for two reasons. First, the evaluation will help the teacher or program developer in making decisions on selecting the appropriate textbook. Furthermore, evaluation of the merits and demerits of a textbook will familiarize the teacher with its probable weaknesses and strengths. This will enable teachers to make appropriate adaptations to the material in their future instruction. In this line, Cunningsworth (1995) and Ellis (1997) propose that textbook evaluation can be of three types, namely 'pre-use', 'in-use', and 'post-use' evaluations. Evaluation of textbooks for pre-use, or predictive, purposes helps teachers in selecting the most appropriate textbook for a given language classroom by considering its prospective performance. The second type of evaluation aids the teacher to explore the weaknesses or strengths of the textbook while it is being used. Finally, post-use, or retrospective evaluation helps the teacher reflect on the quality of the textbook after it has been used in a particular learning-teaching situation.

A checklist is an instrument that helps practitioners in English Language Teaching (ELT) evaluate language teaching materials, like textbooks. It allows a more sophisticated evaluation of the textbook in reference to a set of generalizable evaluative criteria. These checklists may be quantitative or qualitative. Quantitative scales have the merit of allowing an objective evaluation of a given textbook through Likert style rating scales (e.g., Skierso, 1991). Qualitative checklists, on the other hand, often use open-ended questions to elicit subjective information on the quality of course books (e.g., Richards, 2001). While qualitative checklists are capable of an in-depth evaluation of textbooks, quantitative checklists are more reliable instruments and are more convenient to work with, especially when team evaluations are involved.

The review of textbook evaluation checklists within four decades (1970-2000) by Mukundan and Ahour (2010) revealed that most of the checklists are qualitative (e.g., Rahimy, 2007; Driss, 2006; McDonough & Shaw, 2003; Rubdy, 2003; Garinger, 2002; Krug, 2002; McGrath, 2002; Garinger, 2001; Richards, 2001; Zabawa, 2001; Hemsley, 1997; Cunningsworth, 1995; Griffiths, 1995; Cunningsworth & Kusel, 1991; Harmer, 1991; Sheldon, 1988; Breen & Candlin, 1987; Dougill, 1987; Hutchinson & Waters, 1987; Matthews, 1985; Cunningsworth, 1984; Bruder, 1978; Haycraft, 1978; Robinett, 1978); than quantitative (e.g., Canado & Esteban, 2005; Litz, 2005; Miekley, 2005; Harmer, 1998; Peacock, 1997; Ur, 1996; Skierso, 1991; Sheldon, 1988; Grant, 1987; Williams, 1983; Daoud & Celce-Murcia, 1979; Tucker, 1978); or head words/outline format, i.e., those without rating scales or

questions (Ansari & Babaii, 2002; Littlejohn, 1998; Roberts, 1996; Brown, 1995). Most of these checklists are either too short or too long and some criteria in them are vague, so they do not thoroughly meet the requirements of a good and applicable instrument for evaluation purposes.

Despite their crucial roles in language instruction, most if not all the available textbook evaluation checklists have been developed qualitatively often with no empirical evidence in support of their construct validity. Additionally, even when fundamental matters like validity and reliability are accounted for, most of these checklists are impractical. For example, some make use of ELT terminology that sound ambiguous for language instructors with little expertise in the area. A further disadvantage of some of the available checklists is that because of the high number of their items they lack economy and hence practicality. This could be the reason why most language learning materials in the world are evaluated based on the subjective and impressionistic judgment of evaluators.

DETERMINING THE EVALUATIVE CRITERIA

English language teaching (ELT) material developers and evaluators need to take a wide range of factors into consideration before they make decisions on the materials they develop or select for particular contexts. Some of these factors include the roles of the learner, teacher, and instructional materials as well as the syllabus (Richards & Rodgers, 1987). In order to account for these roles effectively, the evaluator must gain an awareness of the learner and teacher's needs and interests (Bell & Gower, 1998).

As it has been argued by some scholars (e.g., Byrd, 2001; Sheldon, 1988), evaluative criteria of checklists should be chosen according to the learning-teaching context and the specific needs of the learner and teacher. However, a review of the available checklists indicates that they have many identical evaluative criteria regardless of the fact that they had been developed in different parts of the world for different learning-teaching situations and purposes. Most well-established checklists such as Cunningsworth and Kusel (1991) or Skierso (1991) examine similar dimensions like physical attributes of textbooks including aims, layout, methodology, and organization. Some other criteria that are present in most checklists include the way language skills (speaking, listening, etc.), sub-skills (grammar, vocabulary, etc.), and functions are presented in the textbook depending on the present socio-cultural setting (Zabawa, 2001; Ur, 1996; Cunningsworth, 1995; Harmer, 1991).

In addition to the criteria mentioned above, a checklist must take into account the background of the target students who are going to use it. The background can encompass a variety of dimensions including students' age, needs and interests (Byrd, 2001; Skierso, 1991). Finally, the language used in the various texts of the textbook under evaluation should present natural and authentic examples of language use in the real world. According to Bell and Gower (1998), employing real language in the textbook contributes to the students' motivation by helping the teacher "get them off the learning plateau" (p. 123). Based on the review of the literature on the textbook evaluation checklists, the researchers created a tentative classification of textbook evaluation criteria (Figure 1).

As the figure shows, we divided the list of criteria into the two general categories including 'general attributes' and 'learning-teaching content'. The first category was further divided into five sub-categories of 'relation to syllabus and curriculum', 'methodology', 'suitability to learners', 'physical and utilitarian attributes', and 'supplementary materials'. The criteria in the second category, on the other hand, included 'general' (i.e., task quality, cultural sensitivity, as well as linguistic and situational realism), 'listening', 'speaking', 'reading', 'writing', 'vocabulary', 'grammar', 'pronunciation', and 'exercises'.

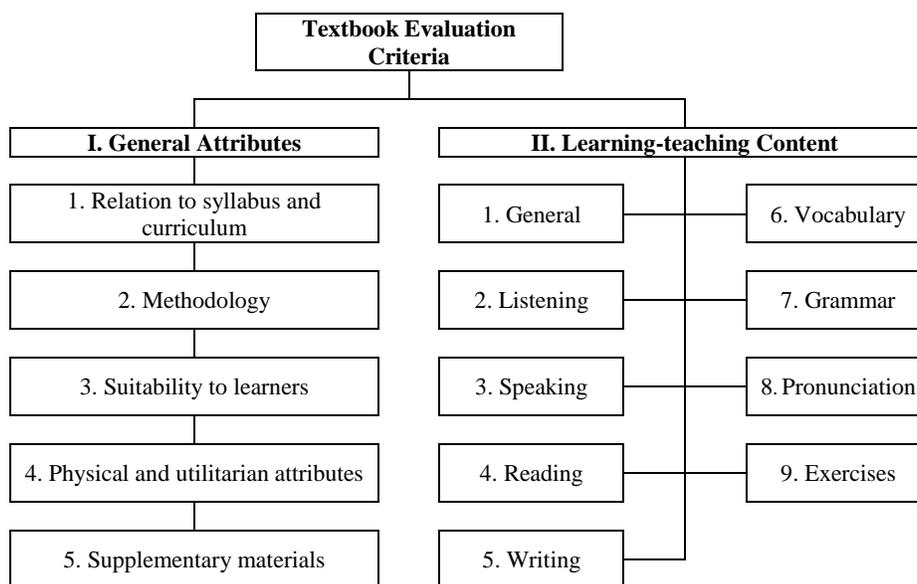


Figure 1. Classification of textbook evaluation criteria

A TENTATIVE CHECKLIST FOR TEXTBOOK EVALUATION

A tentative checklist was developed based on the classification indicated in Figure 1. Appendix (A) presents the checklist which consists of two main sections including ‘general attributes’ and ‘learning-teaching content’. These sections were further divided into several sub-categories following the aforementioned classification. In order to avoid misinterpretations of these features, the developers added one or more descriptors to each sub-category. To offer an example, three descriptors (items I.C.4-6) were added under ‘suitability to learners’. These three items further would indicate a suitable textbook should account for learners’ age, needs and interests.

In the development of this checklist several points had to be considered. First was the issue of validity. To ensure the validity or relevance of any instrument, its developers must be aware of the relevant theories (Messick, 1994). In the light of their present teaching-learning situation, they should consider the construct domain being addressed and specify the criteria for evaluation of ELT materials, which, in turn, will show aspects of the evaluation procedure that may influence the final results and should, therefore, be closely considered. In order to be relevant to the context, a checklist should consider the purpose of evaluation, students, and other features discussed in the preceding section. The present checklist was developed based on a review of the similar previous instruments to ensure its construct validity. Meanwhile, certain items (e.g., items I.A.1, I.C.4-6, II.A.4) in the checklist would bring evaluators’ attention to their present context.

Tomlinson (2003) suggests avoiding large, vague, and dogmatic questions that might be interpreted differently by different evaluators. These factors, if eliminated in the trial process of the developed checklists may result in a more systematic, rigorous, and reliable evaluation. The clarity of the items should, therefore, be taken into consideration. A vague item can decrease the reliability of the instrument. There are certain checklists that fail to elaborate on some items which makes their comprehension very challenging for the novice evaluator. For instance, one of the items in Byrd (2001) describes the ‘Fit between textbook and the curriculum’ as, “fits the pedagogical and SLA philosophy of the program/course” (p.427). Such an item may be easily discernable for an expert in the area; however, it will not be clear enough for an end-user with a low expertise. Developers should seek to design clear items if they really wish their checklist to be utilized. As an example, Skierso (1991) clearly describes the criterion of ‘vocabulary load’, as “the number of new words introduced every lesson” (p. 446). This contributes to the clarity and, in turn, to the reliability of the instrument.

In addition to validity and reliability, checklist developers should also take matters of practicality into account. A checklist must be economical, for instance. Cunningsworth (1995) suggests, “It is important to limit the number of criteria used, the number of questions asked, to manageable proportions, otherwise we risk being swamped in a sea of details” (p. 5). If a checklist is precise and short, it will be expeditious and save considerable time and budget once it is utilized for evaluation purposes. According to Mukundan and Hour (2010), the length of the quantitative textbook evaluation checklists in their study ranged between 113 (Tucker, 1978) and 4553 (Skierso, 1991) running words. The developers of the present checklist made an attempt to come up with a relatively concise instrument. The number of running words in this checklist is 356, which turns out to be moderate compared to other similar checklists. This contributes to its economy.

Another subject that should be noted in the operationalization of a checklist is the numerical value of its items. According to the literature, when faced with an odd-numbered scale, evaluators usually go for a middle score (McColloy & Remsted, 1965). This is known as the problem of “central tendency”; that is, “the inclination to rate people in the middle of the scale even when their performance clearly warrants a substantially higher or lower rating” (Grote, 1996, p. 138). To offer an example, in a five-point scale, an evaluator will more probably assign 3. Therefore, it is advisable to avoid odd-numbered scales when developing an instrument (Sager, 1972). However, as it is the case in most of the available checklists, a rating scale of 0-4 (where 4= Excellent, 3= Good, 2= Adequate, 1= Weak, and 0= Totally lacking) is the dominant form employed (e.g., Daoud & Celce-Murcia, 1979; Skierso, 1991). Furthermore, the end-users of checklists can be informed of the problem of central tendency and be advised to avoid it. Having considered these variables, a scale developer can hope to come up with a fair and viable checklist.

CONCLUSION

This paper discussed the development of a tentative textbook evaluation checklist. For this purpose, the related literature was reviewed and a list of important evaluative criteria was created. This list was further developed by adding one or more items under each category to describe it in detail. A five-point scale was added to the checklist to help the evaluators assess the textbook in reference to each item.

A checklist of this type could be useful for pre-use, in-use and post-use textbook evaluation purposes (Ellis, 1997; Cunningsworth, 1995). Based on the results of such forms of evaluation, substantial educational and administrative decisions could be made that may have financial, professional, and/or political implications (Sheldon, 1988). The checklist could prove informative and useful for curriculum designers, ELT material developers or teachers in the classroom providing them with useful ideas according to which the materials being evaluated can be improved.

The present checklist as it appears in this paper can be further refined through qualitative and/or quantitative studies. Focus group interviews can help the developers improve the clarity of the instrument. Furthermore, a survey of ELT material experts’ evaluation of the checklist and a factor analysis of the collected data can provide empirical evidence for its inclusiveness. Finally, further research is also needed to test the correlation between the findings of the current checklist as compared to those of other well-established instruments.

ABOUT THE AUTHORS

Jayakaran Mukundan is a lecturer at the Faculty of Educational Studies, UPM, Malaysia and a visiting fellow at Leeds Metropolitan University, UK. He is also a Director on the Extensive Reading Foundation Board. His areas of interest include ESL writing and ELT textbook evaluation.

Reza Hajimohammadi holds a Master and a Bachelor degree of TEFL. Prior to pursuing his PhD at University Putra Malaysia (UPM) he has been a lecturer in different universities and schools of Iran since 1989. His areas of interests include language teaching methodology, writing skill, and vocabulary learning.

Vahid Nimehchisalem has been an English teacher, material developer, test developer, teacher trainer and lecturer. His areas of research interest include assessing writing, English Language Teaching (ELT) material evaluation and

ELT methods. He completed his PhD in Teaching English as a Second Language (TESL). He is involved in several research projects in the area of ELT in Universiti Putra Malaysia (UPM).

REFERENCES

1. Ansari, H. & Babaii, E. (2002). Universal Characteristics of EFL/ESL Textbooks: A step towards systematic textbook evaluation. *The Internet TESL Journal*, 8 (2).
2. Bell, J. & Gower, R. (1998). Writing course materials for the world: A great compromise. In B. Tomlinson (ed.). *Material development in language teaching* (pp. 116-129). Cambridge: Cambridge University Press.
3. Breen, M. & Candlin, C. (1987). What materials? A consumer's and designer's guide. In L. E. Shelden (ed.). *ELT textbook and materials: Problems in evaluation and development*. ELT Documents 126. (pp. 13-28) Oxford: Modern English Publications in association with The British Council.
4. Brown, J. D. (1995). *The elements of language curriculum*. New York: Heinle & Heinle Publishers.
5. Bruder, M. N. (1978). Evaluation of foreign language textbooks: A simplified procedure. In H. S. Madsen & J. D. Bowen, eds. *Adaptation in language teaching*. Appendix 2 (pp. 209-218). Rowley, Mass: Newbury.
6. Byrd, P. (2001). Textbooks: Evaluation for selection and analysis for implementation. In M. Celce-Murcia. *Teaching English as a second or foreign language*. (3rd Ed.) (pp. 415-427). US: Heinle & Heinle, Thomson Learning Inc.
7. Canado, M.L.P. & Almagro Esteban, A. (2005). Authenticity in the teaching of ESP: An evaluation proposal. *Scripta Manent* 1(1), 35-43. Retrieved from http://www.sdutsj.edus.si/ScriptaManent/2005_1/Perez_Almagro.pdf
8. Cunningsworth, A. (1984). *Evaluating and selecting EFL teaching materials*. London: Heinemann Educational Books.
9. Cunningsworth, A. (1995). *Choosing your coursebook*. Oxford: Heinemann.
10. Cunningsworth, A., & Kusel, P. (1991). Evaluating teachers' guides. *ELT Journal* 45(2): 128-139.
11. Daoud, A. M. & Celce-Murcia, M. (1979). Selecting and evaluating a textbook. In M. Celce-Murcia & L. McIntosh, eds. *Teaching English as a second or foreign language* (pp. 302-307). New York: Newsbury House.
12. Dougill, J. (1987). Not so obvious. In L. E. Shelden (ed.). *ELT textbook and materials: Problems in evaluation and development*. ELT documents 126. (pp. 29-36) Oxford: Modern English Publications in association with The British Council.
13. Driss, L. (2006). Systematic textbooks supplementation (knowing when to supplement and what to add). US Department of State, Ministry of Education Morocco, University of Delaware, English Language Institute. Retrieved from <http://www.udel.edu/eli/2006P4L/driss.pdf>
14. Ellis, R. (1997). *SLA research and language teaching*. Oxford: Oxford University Press.
15. Garinger, D. (2001). Textbook evaluation. *TEFL Web Journal*. Retrieved from <http://www.teflweb-j.org/v1n1/garinger.html>
16. Garinger, D. (2002). Textbook selection for the ESL classroom. *Center for Applied Linguistics Digest*. Retrieved from <http://www.cal.org/resources/Digest/0210garinger.html>
17. Grant, N. (1987). *Making the most of your textbook*. London: Longman Inc.
18. Griffiths, C. (1995). Evaluating materials for teaching English to adult speakers of other languages. *Forum* 33(3): 50. Retrieved from <http://forum.state.gov/vols/vol33/no3/p50.htm>
19. Grote, R. D. (1996). *The complete guide to performance appraisal*. New York: AMACOM Books.
20. Harmer, J. (1991). *The practice of English language teaching*. Harlow, Essex: Longman.
21. Harmer, J. (1998). *How to teach English*. Harlow, Essex: Addison Wesley Longman.
22. Haycraft, J. (1978). *An introduction to English language teaching*. London: Longman.
23. Hemsley, M. (1997). The evaluation of teachers' guides-design and application. *ELTED* 3(1): 72-83.
24. Hutchinson, T. & Waters, A. (1987). *English for specific purposes: A learning-centered approach*. Cambridge: Cambridge University Press.
25. Krug, K. (2002). Some points to ponder in textbook evaluation. Retrieved from <http://www.kutztown.edu/library/materials/textbookeval.htm>

26. Littlejohn, A. (1998). The analysis of language teaching materials: Inside the Trojan horse. In B. Tomlinson (ed.) *Materials development in language teaching*. Cambridge: Cambridge University Press, 190-216.
27. Litz, D.R.A. (2005). Textbook Evaluation and ELT Management: A South Korean Case Study. *Asian EFL Journal*. Retrieved from http://www.asian-efl-journal.com/Litz_thesis.pdf
28. Matthews, A. (1985). Choosing the best available textbook. In A. Matthews, M. Spratt, and L. Dangerfield (eds.). *At the chalkface*. (pp. 202-206). London: Edward Arnold.
29. McColloy, W. & Remsted, R. (1965). Composition rating scale for general merit: An experimental evaluation. *Journal of Educational Research*, LIX: 55-57.
30. McDonough, J. & Shaw, C. (2003). *Materials and methods in ELT: A teacher's guide*. (2nd Ed.). Malden, MA: Blackwell Publishing Ltd.
31. McGrath, I. (2002). *Materials evaluation and design for language teaching*. Edinburgh: Edinburgh University Press.
32. Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher*, 23(2), 13–23.
33. Miekley, J. (2005). ESL textbook evaluation checklist. *The Reading Matrix*, 5(2). Retrieved from: http://www.readingmatrix.com/reading_projects/miekley/project.pdf
34. Mukundan, J. & Ahour, T. (2010). A review of textbook evaluation checklists across four decades (1970-2008). In Tomlinson, B., Masuhara, H. (Eds.). *Research for materials development in language learning: Evidence for best practice* (pp. 336-352). London: Continuum.
35. Peacock, M. (1997). *Choosing the right book for your class*. Retrieved from <http://www.sx.ac.uk/linguistics/pgt/egspll/volume1/PDFs/PEACOCK1.pdf>
36. Rahimy, R. (2007). A coursebook evaluation. *ESP World*, 6, Issue 2(15). Retrieved from <http://www.esp-world.info/contents.htm>
37. Richards, J. C. (2001). *Curriculum development in language teaching*. Cambridge: Cambridge University Press.
38. Richards, J., & Rodgers, T. (1986). *Approaches and methods in language teaching*. Cambridge: Cambridge University Press.
39. Roberts, J. T. (1996). Demystifying materials evaluation. *System*, 24(3): 375-389.
40. Robinett, W. (1978). *Teaching English to speakers of other languages: Substance and technique*. Minneapolis: University of Minnesota Press.
41. Rubdy, R. (2003). Selection of materials. In B. Tomlinson. *Developing materials for language teaching*. (pp. 37-57). London: Continuum.
42. Sager, C. (1972). Improving the quality of written composition through pupil use of rating scale. PhD Dissertation. Boston: Boston University.
43. Sheldon, L. E. (1988) Evaluating ELT textbooks and materials. *ELT Journal*, 42(4):237-246.
44. Skierso, A. (1991). Textbook selection and evaluation. In M. Celce-Murcia. *Teaching English as a second or foreign language* (pp. 432-453). (2nd Ed.). Boston: Heinle & Heinle Publishers.
45. Tomlinson, B. (2003), Materials evaluation, in B. Tomlinson (ed.), *Developing materials for language teaching*. London: Continuum.
46. Tucker, C. A. (1978). Evaluating beginning textbooks. In H. S. Madsen and J. D. Bowen. *Adaptation in language teaching*. Appendix 3 (pp. 219-237). Rowley, Mass: Newbury.
47. Ur, P. (1996). *A course in language teaching: Practice and theory*. Cambridge: Cambridge University Press.
48. Williams, D. (1983). Developing criteria for textbook evaluation. *ELT Journal* 37(3): 251-255.
49. Zabawa, J. (2001). Criteria for FCE textbook evaluation: an attempt at questionnaire formulation. Retrieved from: http://www.univ.rzeszow.pl/fil_ang/wsar2/sar_v2_17.pdf

Appendix A: A Tentative Checklist For Textbook Evaluation

I. General attributes					
A. The book in relation to syllabus and curriculum					
1. It matches to the specifications of the syllabus.	①	②	③	④	
B. Methodology					
2. The activities can be exploited fully and can embrace the various methodologies in ELT.	①	②	③	④	
3. Activities can work well with methodologies in ELT.	①	②	③	④	
C. Suitability to learners					
4. It is compatible to the age of the learners.	①	②	③	④	
5. It is compatible to the needs of the learners.	①	②	③	④	
6. It is compatible to the interests of the learners.	①	②	③	④	
D. Physical and utilitarian attributes					
7. Its layout is attractive.	①	②	③	④	
8. It indicates efficient use of text and visuals.	①	②	③	④	
9. It is durable.	①	②	③	④	
10. It is cost-effective.	①	②	③	④	
E. Efficient outlay of supplementary materials					
11. The book is supported efficiently by essentials like audio-materials.	①	②	③	④	
II. Learning-teaching content					
A. General					
1. Most of the tasks in the book are interesting.	①	②	③	④	
2. Tasks move from simple to complex.	①	②	③	④	
3. Task objectives are achievable.	①	②	③	④	
4. Cultural sensitivities have been considered.	①	②	③	④	
5. The language in the textbook is natural and real.	①	②	③	④	
6. The situations created in the dialogues sound natural and real.	①	②	③	④	
B. Listening					
7. The book has appropriate listening tasks with well-defined goals.	①	②	③	④	
8. Tasks are efficiently graded according to complexity.	①	②	③	④	
9. Tasks are authentic or close to real language situations.	①	②	③	④	
C. Speaking					
10. Activities are developed to initiate meaningful communication.	①	②	③	④	
11. Activities are balanced between individual response, pair work and group work.	①	②	③	④	
D. Reading					
12. Texts are graded.	①	②	③	④	
13. Texts are interesting.	①	②	③	④	
E. Writing					
14. Tasks have achievable goals and take into consideration learner capabilities.	①	②	③	④	
15. Tasks are interesting.	①	②	③	④	
F. Vocabulary					
16. The load (number of new words in each lesson) is appropriate to the level.	①	②	③	④	
17. There is a good distribution (simple to complex) of vocabulary load across chapters and the whole book.	①	②	③	④	
18. Words are efficiently repeated and recycled across the book.	①	②	③	④	
G. Grammar					
19. The spread of grammar is achievable.	①	②	③	④	
20. The grammar is contextualized.	①	②	③	④	
21. Examples are interesting.	①	②	③	④	
22. Grammar is introduced explicitly and reworked incidentally throughout the book.	①	②	③	④	
H. Pronunciation					
23. It is contextualized.	①	②	③	④	
24. It is learner-friendly with no complex charts.	①	②	③	④	
I. Exercises					
25. They are learner friendly.	①	②	③	④	
26. They are adequate.	①	②	③	④	
27. They help students who are under/over-achievers.	①	②	③	④	

NOTES