Title:
Development of a New Checklist for Evaluating Reading Comprehension Textbooks

Authors:

Fatemeh Mahbod Karamoozian

Abdolmehdi Riazi, Ph.D

Bio Data:
Abdolmehdi Riazi holds a Ph.D. in TESL from the OISE/University of Toronto. He is currently an Associate Professor in the Department of Foreign Languages and Linguistics of Shiraz University. His areas of interest include academic writing, language learning strategies and styles, and language assessment.

Email address: ariazi@shirazu.ac.ir

Fatemeh Mahbod Karamoozian is an MA student in TEFL at Azad University of Bandar Abbas. She is currently writing her MA thesis and teaching English courses in different institutes in Kerman.

Email address: kkaramoozian@yahoo.com

Qualifications:
Abdolmehdi Riazi holds a Ph.D. in TESL from University of Toronto, Canada.
Fatemeh Mahbod Karamoozian is an MA student in TEFL at Azad University of Bandar Abbas, Iran.

2008

Abstract

The purpose of the present study was to explore the main quality features of several available textbook evaluation checklists proposed by researchers and professionals in this field while it has a special focus on reading comprehension textbook evaluation schemes and checklists. All checklists were explored in terms of their format, scope, applied terms and concepts, weighting/rating systems, guidelines, and whether they were piloted by the developers. In the field of textbook evaluation research, methods are rarely discussed clearly and in-depth. However, three basic methods of textbook evaluation can be discerned in the related literature: impressionistic method, checklist method, and in-depth method. Compared to the two other alternatives, impressionistic evaluation and in-depth evaluation, the checklist method is known to be more advantageous. The results of this study revealed that although the reviewed checklists have several strong points specifically regarding their format and scope, they mostly fail in terms of other features that lead to practicality. Furthermore, the available checklists are mostly designed to evaluate general English textbooks while they are not generalizable enough to be adapted to evaluate other English language textbooks. This study also focused on several specifically developed checklists for evaluating reading comprehension textbooks, while the results revealed that they also suffer from the same shortcomings. Therefore, based on the current study, the authors have developed a new checklist in which they have eliminated the mentioned defects. This checklist is a comprehensive reference specifically for evaluating reading comprehension textbooks while flexible enough to be used for other purposes too.

Key words: textbook evaluation, textbook evaluation schemes/frameworks, textbook evaluation checklists, reading comprehension

Development of a New Checklist for Evaluating Reading Comprehension Textbooks

Introduction

In the field of textbook evaluation research, methods are rarely discussed clearly and in-depth. Taking into account research procedures, data processing, and interpretation; textbook evaluation methods can be classified into a very detailed list: a) methods of theoretical analysis: 1) the theoretical-analytical methods (e.g. the determination of the conformity between the textbook and the syllabus - comparative study), 2) the special analytic method (i.e. analysis according to a set of internal didactic criteria), and 3) the comparative analysis of textbook (i.e. two or more textbooks are mutually compared); b) empirical-analytical methods: 1) experimental investigation in the use of textbooks, 2) public inquiry applied to teachers, 3) public inquiry applied to learners, and c) Statistical (quantitative) methods (Hrehovcik, 2002).

However, from a simpler perspective, only three basic methods of textbook evaluation can be discerned in the related literature: impressionistic method, checklist method, and in-depth method. The impressionistic method is concerned to obtain a general impression of the material and involves glancing at the publisher's blurb and content pages of each textbook, and then skimming throughout the book looking at various features of it. In-depth techniques go beneath the publisher's and author's claims. It considers the kind of language description, underlying assumptions about learning or values on which the materials are based or, in a broader sense, whether the materials seem likely to live up to the claims that are being made for them (McGrath, 2002).

The checklist method contrasts system (objectivity) with impression (subjectivity). Compared to the two other alternatives, impressionistic evaluation and in-depth evaluation, the checklist has at least four advantages: it is systematic which ensures that all elements that are deemed to be important are considered, it is cost effective which permits a good deal of information to be recorded in a relatively short space of time; the information is recorded in a convenient format which allows for easy comparison between competing sets of material; and it is explicit, and, provides the categories that are well understood by all involved in the evaluation while offers a common framework for decision making (McGrath, 2002).

The checklist method is advocated by most experts. For instance, Tomlinson (1998) supports the use of this method and maintains that one of the most obvious sources for guidance in analyzing materials is the large number of frameworks which exist to aim in the evaluation of a textbook. However, as he mentions the checklist typically contains implicit assumptions about what desirable materials should look like, and each of these areas might be debatable while also limit their applicability.

Objectives of the Current Study

This study attempts to explore the main quality features of several available textbook evaluation checklists, proposed by researchers and professionals in this field while it has a special focus on reading comprehension textbook evaluation schemes and checklists. All checklists are investigated in terms of their strong and weak points and their practical limitations. The results of this study will provide insights towards the most valuable features that should be considered in a checklist which might affect its practicality. This could be a matter of concern to researchers, reviewers, administrators, educators, educational advisors, textbook and checklist developers, and publishers.

A review of some checklists, their strong and weak points

Rivers (1968) presented a set of criteria for textbook evaluation which is based on seven major areas: 'appropriateness for local situation' (including 6 evaluative items), 'appropriateness for teachers and students' (10 items), 'language and ideational content' (6 items), 'linguistic coverage and organization' (9 items), 'types of activities' (6 items), 'practical considerations' (7 items), and 'enjoyment index' (1 item). Each of these evaluating items also includes several questions that aim at evaluating some features. The rating system in River's checklist (pp.477-483) is based on a 5-point scale: excellent for my purpose (1), suitable (2), will do (3), not very suitable (4), and useless for my purpose (5).

Rivers' checklist seems to focus both on several detailed and major points. For instance, item 14 denotes some detailed points which are ignored in most other checklists: "is there a table of contents setting out clearly which structures are introduced and in what order?" (p.479), while items like 23 denote some major points: "how is pronunciation dealt with?" (p.480). One valuable point about the checklist is that it makes use of clear terms and if needed it presents enough explanations on several concepts (e.g., item 37: "is some material introduced just for fun and relaxation…"). Considering its comprehensiveness, this checklist can be a good source of ideas or a reference about what points to consider in developing checklists or writing materials. However, with regard to its practicality in evaluating textbooks several defects exist. One point is that the checklist is not presented in a user-friendly format. The other point is that the user is faced with some main areas that should be rated, while each of these areas include several other items and features that cover a variety of points. Therefore, the user should consider several features and assign just one rate as a total, while the author has not provided any guidance on how to perform it. Also, according to this rating system it is not really possible to compare several competing textbooks, as well as each area of one book. Another important point is that the weighting system is based on the use of some terms including 'excellent for my purposes' to 'useless for my purpose'. These terms are just indicators of the suitability of the materials for a special context but not the appropriacy of them and this might cause variety of answers even in a same situation.

Daoud and Celce-Murcia (1979 cited in Celce-Murcia, 2001) introduced a broad evaluative checklist. They consider five major components for the textbook in their checklist: subject matter (including 4 evaluative items), vocabulary and

structures (9 items), exercises (5 items), illustrations (3 items), and physical make-up (4 items). Also there is a section for teacher's manual which includes four major parts: general features (including 5 evaluative items), type and amount of supplementary exercises for each language skill (6 items), methodological/pedagogical guidance (7 items), and linguistic background information (4 items). The rating system is based on a 5-point scale: excellent (4), good (3), adequate (2), weak (1) and totally lacking (0).

Daoud and Celce-Murcia's checklist seems to be organized while it categorizes some main and familiar concepts like exercises, illustrations, etc. The format of the checklist is user-friendly and seemingly easy to follow. Therefore, the format can be a reference in designing checklists. Furthermore, the checklist pays enough attention to the existence and content of teacher's manuals that are not taken seriously in most other checklists. However, when the user starts working on the items, s/he will face several difficulties. One of the major problems with this checklist is that although the terms which are used in each item are simple, but more explanation on several items is needed to enable the user to apply them according to the intention of the authors. For instance, in vocabulary and structure section, item 2 is: "are the vocabulary items controlled to ensure systematic gradation from simple to complex items" (p.425)? To weight this item, however, one needs to know how the authors define simple and complex items. In the same section, items 4 and 5 are: "does the sentence length seem reasonable for the students of that level?" and "is the number of grammatical points as well as their sequence appropriate" (p.425)? In item 4 a criterion for the relation between sentence length and the level of students is not provided. This is also true about the number of grammatical points and their sequence, in item 5. The existence of such items in a checklist causes some serious problems: the user may ignore them or answer them inattentively, or s/he may consume more time to find a criterion that helps her/him in answering them. While this criterion might not be reliable or might be achieved from various sources that cause variety of answers even in the same situation. For the items that include more than one feature such as "do illustrations create a favorable atmosphere for practice in reading and spelling by depicting realism and action?" the user is not provided with any guidance on how to rate them (p.425). The last point is that the checklist is not piloted by the authors and/or the authors have not provided an organized set of guidelines that facilitate its use.

Williams (1983) presented a scheme for evaluating ESL/EFL textbooks with the claim that it could be adapted for particular contexts. The scheme includes these features: up-to-date methodology of L2 teaching, guidance for non-native speakers of English, needs of learners, and relevance to socio-cultural environment. Each of these features can be evaluated in terms of linguistic/pedagogical aspects: general, speech, grammar, vocabulary, reading, writing, and technical. For each of these aspects, then, four evaluative items are considered to provide a checklist. The weighting system of this checklist is based on a 5-point scale: 0-4.

Williams' checklist seems to have a user-friendly and easy to follow format. The scheme mainly focuses on very broad quality features of textbook and concepts of teaching/learning language while it really ignores some important and detailed points. Moreover, the author has not provided enough explanation to clarify the concepts of each item. The checklist has tried to consider local differences and first language effects important in language learning; for instance, in general section; items

3 and 4: "caters for individual differences in home language background" and "relates content to the learners' culture and environment" (p.255). There are several of such items in other sections, too. As the checklist does not take into account detailed features one might find less difficulties answering items such as "selects vocabulary on the basis of frequency, functional load, etc." (p.255). The reason is that the user can answer this item by referring to the information provided by the author or publisher in the introduction or appendices of the book and rate the item. However, this information might be inadequate in terms of several detailed features which affect vocabulary selection such as target students' needs and levels. As the most items in this checklist include such broad concepts, the problem is that by using these types of checklists, the value of one textbook might be underestimated or overestimated. The same as most other checklists William's checklist is not also piloted by the author and there is no guideline provided for it as how to use or adapt its content.

Sheldon (1988) presented a checklist that includes two main categories: factual details and factors. Factual details contain the title, author, publisher, price, physical size, duration of the course, target learner, teacher, and skill. Factors include rationale, availability, user definition, layout/graphics, accessibility, linkage, selection/grading, physical characteristics, appropriacy, authenticity, sufficiency, cultural bias, educational validity, stimulus/practical revision, flexibility, guidance, and overall value for money. Each of the factors includes 3, 2, 3, 2, 4, 3, 3, 4, 3, 3, 2, 6, 1, 3, 3, 6, and 2 evaluative items. The assessment in this checklist is based on a 4-point scale: poor, fair, good, and excellent.

Sheldon's checklist focuses both on detailed and major points including quality features of textbooks and theories of teaching/learning. For instance, items such as "can you contact the publisher's representatives in case you want further information about the content, approach, or pedagogical detail of the book?"(p.243) denote some detailed points which are ignored in most other checklists. While some other items such as "does the textbook cohere both internally and externally (e.g., with other books in a series)?" (243) refer to some broad points. Considering its comprehensiveness, this checklist can be also a good source of ideas or a reference about what points to consider in developing checklists or writing materials. However, with regard to its practicality in evaluating textbooks several problems occur. The existence of some complicated terms is one major problem with this checklist. For example, "is it pitched at the right level of maturity and language, and…at the right conceptual level?"(p.244). It is not clear what the author means by 'right level of maturity' and 'right conceptual level.' Some items need more explanation as they are ambiguous or they need a criterion to be followed. For instance, " does the introduction, practice, and recycling of new linguistic items seem to be shallow/steep enough for your students" (p.243) or "..are the texts unacceptably simplified or artificial (for instance, in the use of whole-sentence dialogues)?" (p.244). In these items the criterion for being 'shallow/steep enough' or 'unacceptably simplified or artificial' is not defined, while the terms might denote several aspects. Moreover, it is not clear that each item or feature of the checklist should be considered in a whole book or for each unit/lesson of it. One other difficulty is the scale, which makes use of terms and asterisks. Therefore, for those items that include several features how it would be possible to consider just one rate as a total, while the rating system does not allow rating each item/feature and calculating an average for the whole.

Furthermore, according to this rating system it is not really possible to compare several competing textbooks, while the user can not make an average of the rates of the whole checklist as well as each section of it. Another point is that the weighting system is subjective as one should rate each area from 'poor' to 'excellent' and this may cause variety of answers even in a same situation. Furthermore, as the weighting is solely based on some terms (poor, fair, good, and excellent), the results can not be graphically displayed. The last point that worth considering is that the author has not piloted his checklist and/or has not provided the users with a guideline on how to apply or adapt it.

Griffiths (1995) presented a list of questions as criteria for evaluating materials for use with speakers of other languages. These questions examine 12 features of materials: the match between material and learner objectives, learner-centered material, facilitating interactive learning, socio-cultural appropriateness, gender sensitivity, up-to-date materials, well-graded vocabulary and comprehensible input, age-appropriate materials, interesting and visually attractive material, relevance to real life, easy to use material, and ethnocentric material.

Griffiths' checklist makes use of clear terms and provides users with a short explanation for each item. This can be really worthwhile as it helps the checklist users to understand the rationale behind each item and reveals the major points that the author intended. However, some problems exist with this checklist. The checklist does not introduce any criteria on which items such as "are vocabulary and comprehensive input levels well-graded?" (p.50) can be judged. Moreover, the items are too broad and several important points are ignored (e.g., language items and skills) which might lead to underestimation or overestimation. The items are open-ended questions; therefore, the weighting system is totally subjective, the comparison of various textbooks or various sections of a textbook is not possible, and the results of the evaluation can not be displayed graphically. Moreover, the checklist lacks any guideline and is not piloted by the author.

Ur (1996) presented a set of general criteria for assessing any language-teaching textbooks which includes a list of several criteria composed of nineteen features.These features include: objectives being explicitly laid out in an introduction and implemented in the material, approach educationally and socially to the target community, clear attractive layout and easy print to read, appropriate visual materials available, interesting topics and tasks, varied topics and tasks, clear instructions, systematic coverage of syllabus, clearly organized and graded content, periodic review and test sections, plenty of authentic language, good pronunciation, vocabulary and grammar explanation and practice, fluency practice in all four skills, encouraging learners to develop their own learning strategies and to become independent, adequate guidance for teacher; audio cassettes, and being readily available locally. The rating is based on a 5-point scale: a double tick (very important), a single tick (fairly important), a question mark (not sure), a cross (not important) and a double cross (totally unimportant) are used to rate the items.

The format of Ur's checklist is user-friendly, seemingly easy to follow, and it contains clear terms. However, when the user starts working on it, s/he will face several difficulties. Although the checklist takes into consideration several important

detailed points such as "clear instructions" or "periodic review and text sections" (p.186), it mostly contains very broad concepts such as "good pronunciation explanation and practice" and "fluency practice in all four skills" (p.186). Therefore, the evaluation might reveal the results that are overestimated or underestimated. Moreover, the criteria according to which these items should be evaluated are not provided. Another problem about this checklist is that it is not balanced in terms of the features it considers about physical aspects of a book, language items, skills, etc. The rating system including a range of 'very important' to 'totally unimportant' has also several previously mentioned defects. It limits comparison of several textbooks or sections of a book, limits graphic displays, and for items that include more than one evaluating feature, it is not possible to assign one rate.

Peacock (1997) presented a more objective checklist. As he argued, while it can correspond to local needs it is flexible enough to be used worldwide, and is designed to evaluate EFL textbooks from beginning to upper intermediate adult learners. The goal of the checklist, as he mentions, is not to analyze textbooks in great depth from a linguistic or pedagogic viewpoint, but to allow as thorough an evaluation as possible to be made in the time normally allocated for textbook assessment by EFL teachers. Peacock's checklist contains eight sections: general impression, technical quality, cultural differences, appropriacy, motivation and the learner, pedagogic analysis, finding the way through the student's book and supplementary materials. Each of these sections includes 3, 3, 3, 4, 7, 25, 8 and 7 items respectively. The checklist is based on a scoring table with weightings that can be varied by users according to any local situation. The rating system is based on a 3-point scale: good (2), satisfactory (1) and poor (0).

Peacock's checklist begins with somehow the same part as 'factual details' of Sheldon's (1988). The checklist consists of several sections that make it more organized and easy to follow. Moreover, the author has provided an instruction for his checklist, as for which books it will work and also he has considered the importance of comparing several textbooks, while rating them. The texts and the terms of the checklist are clear enough. The checklist is accompanied by a scoring table, in which the users are asked to first weight each item according to their local situation and the importance that each item might have there. Then they are asked to multiply the score of each item by the weighting value. This system of rating seems to open a room for more considerations about the importance of learners and teachers' needs in variety of situations. Also a number of items in the checklist are considered to be rated according to local settings and target learners. The checklist focuses on variety of

detailed and basic concepts, therefore seems to be comprehensive. For instance, item 13 and 18: "the activities are adaptable to personal learning and teaching styles" and "the book encourages learners to assume responsibility for their own learning" (p.7), refer to important qualities that are not considered in most checklists while items such as 22: "methodologically the book is in line with current worldwide theories and practices of language learning" (p.7) focuses on a broad concept. One more point is that although the checklist really seems to be more practical than some other ones, it still suffers from major defects. As most checklists the rating system is not reliable enough, and there is no advice on how to rate the items that include more than one evaluative feature. Furthermore, the checklist is classified into some sections but the scoring table ignores this point and just considers individual rates, while it might again limit the comparison of several sections of a book or some books. Another point is that for some items such as 36 and 37 the criteria are not defined: "In general the activities in the book are neither too difficult nor too easy for your learners" (p.7) and "the book is sufficiently challenging to learners" (p.8). Like most other checklists it is not accompanied by guidelines, however, it is piloted which is a positive point for this checklist.

Harmer (1998) states that there are nine main areas which teachers should consider in the books they evaluate: price, availability, layout and design, methodology, skills, syllabus, topic, stereotyping, and the teacher's guide. He considers 5, 6, 5, 3, 5, 4, 5, 4 and 5 evaluative items for each of these areas, respectively. The weighting in this checklist is based on the descriptive answers provided by the users.

Harmer's checklist consists of several definite sections. The texts and the applied terms are easy and comprehensible. The checklist considers some detailed points such as: "how user-friendly is the design" (p.119), but it really fails to consider several important and basic qualities such as language items. Moreover, the problem with some items is that the criteria that can be a base of judgement are not defined for them: "do the reading and listening texts increase in difficulty as the book progresses" (p.119). Another problem with this checklist is that it contains items which are open-ended questions, and this makes the base of evaluation subjective. Furthermore, it is limited in comparing several books or sections of the same book, and also presenting graphic displays of the results. One other important point is that as the checklist mostly consists of very broad items, overestimation and underestimation might also happen. As in most other checklists, no guideline is provided and it is not piloted.

Zabawa (2001) suggests a checklist of criteria for the Cambridge First Certificate in English (FCE) textbooks that he argues will work for all EFL textbook. This checklist considers 10 categories: layout and design, material organization, language proficiency, teaching reading comprehension, teaching writing, teaching grammar and vocabulary, teaching listening comprehension, teaching oral skills,

content, and exam practice. These categories include 4, 5, 4, 3, 4, 5, 5, 5, 4, and 5 evaluative items respectively. Rating in this checklist is based on a 5-point scale: unsatisfactory (1), poor (2), satisfactory (3), good (4), and very good (5).

Zabawa's checklist consists of several definite sections. The texts and the applied terms are comprehensible while the author has provided short explanations as rationale for the main criteria he considers. These explanations clarify the importance of each criterion and also help users in a way they focus on main features that the author intended. For instance, he maintains that "by language proficiency I mean both the level of language at the beginning of the course and the progression of the material presented" (p.163). One important point is that to avoid the broad concepts the author has provided several detailed questions for each main evaluating item that reveals the nature of the item and the features that should be weighted. Furthermore, the author considers some valuable points such as "does the textbook develop reading skills and strategies (e.g. skimming and scanning), and not just the ability to answer reading comprehension questions?" in his checklist (p.166). However, while applying this checklist some same defects occur. With regard to the scoring system, as scores should be assigned just to the ten main categories, the author has not provided the user with any additional information about how to consider several evaluating features and assign one final rate. Moreover, while the author has tried to draw attention on some points as the features that should be considered in weighting each item, but they are not organized in a way the user feels the need to follow them. This causes subjectivity in evaluation, while it could be a base of more objective evaluation type. One more point is that although the author has tried to provide a balance among the major criteria specifically those that denote the language items and skills, it seems that some main points such as advice about teaching and learning are ignored. This checklist, the same as most others, is not piloted by the author.

Robinett (1978 cited in Brown, 2001) introduced another checklist. The main categories of this checklist are as follows: goals of the course, background of the students, approach, language skills, general content, quality of practice material, sequencing, vocabulary, general sociolinguistic factors, format, accompanying materials and teacher's guide. Each of these twelve categories also includes 1, 4, 2, 4, 4, 5, 4, 3, 2, 7, 4 and 4 evaluative items respectively.

Robinett's checklist seems to be organized and easy to follow. The text is simple enough and each of the categories contains several features. The evaluative items weight the features that are generally broad concepts and also some detailed ones. For instance, 'does the book pay sufficient attention to strategies for word analysis' refers to a detailed point and 'does the theoretical approach reflected in the book reflect a philosophy that you and your institution and your students can easily identify with?' refers to a broad concept. However, some important points are missed in the checklist for example concepts related to teaching and learning language items and skills. This might lead to overestimated or underestimated evaluations. The checklist also suffers from some more shortcomings. No systematic rating scale is discerned therefore the items might be answered in a totally subjective way, and the comparison and graphical displays of the results will not be possible. In this checklist also several items need more explanation and criteria that are defined by the author to be followed. For instance, the following item: "appropriateness and currency of topics, situations, and contexts" (p.142) needs more explanation. This checklist is not accompanied by a guideline and is not piloted by the author either.

Byrd (2001 cited in Celce-Murcia, 2001) developed a checklist that includes 4 main evaluative categories: the fit between the textbook and the curriculum, the fit between the textbook and the students, the fit between the textbook and the teachers, and overall evaluation of the fit of the book for the course in the program. Each of these categories also includes some sub-categories: 4 sub-categories in category 1, 4 in 2, 6 in 3, and 1 in 4. The system of rating is based on a 4-point scale: yes (a good fit), perhaps (an adequate fit), probably not (a poor fit), and absolutely not (wrong for curriculum, students, and/or teachers).

Byrd's checklist seems to be organized and user-friendly but it merely includes some major points and ignores several important and detailed features. Moreover, such broad and short items lead to ambiguity and therefore overestimated or underestimated evaluations. Also, several concepts are not comprehensible and need more explanation or a set of defined criteria to be followed. For instance: "has appropriate linguistic content" or "activities appropriate for students" might be interpreted in different ways (p.427). Anyhow, the items can provide the checklist designers with major points to be considered in materials. The weighting in this

checklist is just based on some terms ('yes' to 'absolutely not'), and the system of evaluation is subjective. The application of such terms limits the comparison among several textbooks. It is not also possible to weight the item which contains several features and assign one term as a whole rate. Moreover, the results cannot be shown graphically.

Garinger (2001) in his study created a checklist for evaluating textbooks in community-based (ESL) programs in a local setting. He considered the most salient features of previously published lists and chose the areas of commonality among them. The features of this checklist are divided into two main categories: practical considerations and language related considerations. Practical considerations include: value/availability, layout/physical characteristics and cultural components. Language related considerations include: skills, language, exercises and user definition. These 7 categories also include 3, 2, 3, 2, 4, 4, and 3 sub-categories respectively. The weighting in this scheme is based on descriptions provided by evaluators. Garinger (2002) presented a scheme in which he considered the intervening options in the process of textbook selection. These options move from broader issues such as program options (like goals and curriculum) to more specific ones (such as exercises and activities). He claims that this scheme enables evaluators to eliminate unsatisfactory textbooks at each stage of analysis and therefore the most appropriate are left at the end. He also presents this scheme in the format of a checklist which includes 4 sections that are: program and course, skills, exercises and activities, and practical concerns that include 6, 3, 4, and 3 items; respectively. Evaluation in this checklist is based on two options of yes and no.

Garinger's (2001) checklist is organized and distinguishes some main items that are classified to more detailed ones. Applied terms are simple, and the checklist mostly contains broad concepts and some detailed ones. For example: "effective use of headings" is detailed while "language grading/sequencing" denotes to a broad concept. The existence of such broad concepts, and the format of the checklist that is shortened by the author, causes that the user feels in need of more explanation or a set of criteria to be followed in evaluation. Also, as the checklist fails in weighting several main points such as language items, the danger of overestimation and underestimation exist. Moreover, the rating system seems to be totally subjective as

the items are open-ended questions. Using this system of rating, the ability to compare several textbooks or several characteristics of a book, and graphic displays, are limited. Garinger's (2002) checklist is more organized than his previous checklist (Garinger, 2001). He has provided some explanations for each main evaluating category that enable checklist users to learn more about the rationale and evaluating features they should consider. Although the checklist focuses on several important features that were missed in his previous checklist, for instance "does the textbook provide learners with adequate guidance as they are acquiring these skills?" (p.4) but it also misses some valuable evaluating features such as cultural factors or language items. Therefore, the danger of overestimated and underestimated evaluations exists. Moreover, the author has tried to facilitate the process of rating items by changing open-ended questions to yes/no questions, but the problems with rating system still remain unsolved. Both checklists are not piloted by the author and while the second one is accompanied by several fine explanations but it is not organized in the form of a guideline.

Ansary and Babaii (2002) argued that most checklists created by authorities have had little practicality. In a survey they used ten EFL/ESL textbook reviews and ten EFL/ESL textbook evaluation checklists and attempted to discover what authors often consider as the important elements of EFL/ESL textbooks. They then selected a set of common characteristics of these textbooks and introduced a universal and broad textbook evaluation scheme. These characteristics include approach, content presentation, physical make-up and administrative concerns. Each of these characteristics also includes 1, 3, 5 and 3 evaluative items while they also contain a number of features that are totally 28. Evaluation in this checklist is based on a Perfect Value Score (PVS) of 2 and a Merit Score (MS) consisting of numbers 0-2 which indicates a total lack to a perfect match.

Ansary and Babaii's checklist is somehow the same as Garinger's (2001) checklist in format. It is organized as it categorized several sections, and seems to be easy to use. However, getting started working on it, several major problems occur. One of these problems is that it is too shortened and this causes a real ambiguity in answering items. For instance, 'selection and its rationale' item, is made up of several features: 'coverage, grading, organization, and sequencing' while it is not clear what

factors do the authors consider for each feature, how each of them should be considered in the book or each part of it, and how they should be rated. Moreover, some terms and the way they are applied, make the concepts complicated: 'dissemination of a vision (theory or approach) about the nature of language, the nature of learning, and how the theory can be put to applied use.' This item denotes what the users will not be able to understand and apply; therefore, more explanation or defined criteria is needed. One more point is that although the checklist includes important detailed points such as 'appropriate title', it really fails to consider several valuable points about language items and skills. Therefore, the danger of overestimation and underestimation exists. The rating system is based on Tucker's (1975) scheme in which the user assigns an ideal rate to each defined criterion and then a comparative weight is assigned to the relative realization in the textbook under scrutiny of each criterion that indicates a match between the ideal defined criterion and its actual realization in a particular textbook (2 for a perfect match, and 0 for a total lack). Then it is suggested that the numbers be represented on a graph by drawing a dotted line corresponding to the numerical value of merit score and a solid line to represent the perfect value score. This system of rating seems plausible as it tries to both consider local importance of some evaluative items and weight them, and also tries to make comparison and contrasts of several textbooks possible. The same as most other checklists no guideline is provided for it, but it is piloted.

Litz (2005) developed a series of textbook evaluation questionnaires which are created to be answered by the instructors and students. The questions included points about: a) practical considerations, b) layout and design, c) activities, d) skills e) language type, f) subject and content and g) conclusion/overall consensus. Each of these categories also contains some evaluative items that are presented respectively for the student and teacher evaluation forms: a) 2 and 5, b) 2 and 8, c) 5 and 7, d) 3 and 5, e) 6 and 6, f) 5 and 5 and g) 2 and 4. The rating system is based on a 10-point scale which moves from highly disagree (1) to highly agree (10).

Litz's checklist is organized and distinguishes several main categories and some detailed items. The items contain simple terms and are mostly comprehensible and the checklist seems to be comprehensive and balanced in terms of various aspects it covers though it lacks some valuable points as instruction on language learning

strategies or advice about using the book. One important point is that the checklist is piloted at the same time therefore the descriptions provided by the author might be valuable in the sense that they give the user some hints about what features to consider. One other point is that although the users are provided with such information, in terms of application they do not know how to judge and where to refer for a source of criteria. For instance, item 15 in student's checklist and 28 in teacher's checklist is: 'the progression of grammar points and vocabulary items is appropriate'. To rate this item, in terms of vocabulary for instance, the author points out some features: frequency, coverage, range, availability, and potential learnability. However, when the user begins weighting the item, s/he is not able to judge the appropriateness of each as s/he is not suggested about what criteria to follow. The rating system in this scheme is based on highly disagree (1) to highly agree (10). This system of rating is the same as most other checklists that do not provide opportunities for comparison and seems to be subjective. Furthermore, the author has not provided any advice on how to rate those items that include more than 1 evaluating feature.

Miekely (2005) provided a tool for evaluating EFL/ESL reading textbooks. He created a checklist based on recent research in language instruction and twenty two previously developed textbook evaluation checklists. He divided the checklist into three major parts including features of textbooks, teacher's manuals and contexts. He considered content, vocabulary and grammar, exercise and activities and attractiveness of the text and physical make-up as the main features of textbooks. Each of these features includes 5, 5, 7 and 4 evaluative items. General features, back ground information, methodological guidance, supplementary exercises and materials are the main features of teacher's manuals, as he suggests. Each of these features includes 2, 2, 3 and 3 evaluative items. For the last part, contexts, he considers appropriateness of the textbook for the curriculum, for the students that will use it and for the teacher that will teach it. These features are also divided into 1, 4 and 1 evaluative items. The system of weighting is possible by assigning options of Mandatory (M), Optional (O), and Not applicable (N) to each item. Furthermore, there is a numerical scale which ranges between 0-4: excellent (4), good (3), adequate (2), poor (1), and totally lacking (0).

Miekely's checklist is specifically designed for evaluating reading comprehension textbooks. It is organized and contains several main categories and some items in each category. The checklist seems to be comprehensive as it contains variety of evaluating points and there is a good balance in presenting them. However, some valuable points such as features that help learners achieve self-dependence in using materials are missed (e.g., instruction on a number of reading comprehension strategies). This checklist also has the problem of ambiguity in some applied terms and several items of it need more explanation or a set of criteria. For instance, in this item: "are the grammar rules presented in a logical manner and in increasing order of difficulty?" (p.4) the criterion of judgment is not described. The rating system also has the same defects as most other checklists. It is based on excellent (4) to totally lacking (0) with some additional options: mandatory, optional, and not applicable that is perhaps added to consider local differences. Anyhow, the rating systems is again subjective, it doesn't allow comparison of several textbooks, and advice on rating several items that contain more than one evaluating feature is not provided (e.g., "are the texts selections representative of the variety of literary genres, and do they contain multiple sentence structures?" (p.4)). The checklist is not piloted and no guideline is provided that help users how to use or adapt and what to weight.

Jahangard (2007) closely examined 10 checklists proposed by different authors and selected 13 features which were common to most of these checklists. The items of his framework are presented through a set of questions containing these features: explicit objectives in the layout and introduction that are implemented in materials, good vocabulary explanation and practice, educationally and socially acceptable approaches, periodic reviews and test sections, appropriate visual materials, interesting topics and tasks, clear instructions, clear attractive layout and print easy to read, content clearly organized and graded, plenty of authentic language, good grammar presentation and practice, fluency practice in all four skills, and developing learning strategies of learners to become independent. The weighting system in this scheme is descriptive.

Jahangard has piloted his checklist on several textbooks. Therefore, the user can find his way through the explanations he has provided for his checklist, while piloting it. However, the suggested features that should be considered are not

organized. Therefore, the user might not feel at ease or find any need to follow them. Most items focus on valuable points but they are generally representative of some broad concepts such as "fluency practice in all four skills" (p.14). One important point is that although several good and detailed points are also included such as "clear instruction" (p.12), several basic features such as advice on using textbooks are missed. The weighting system is based on answers provided by the users therefore it is not possible to compare several textbooks or parts of a book, and/or display the results graphically.

Rahimy (2007) used a two stage evaluation framework which is based on the macro and the micro evaluations. Macro evaluation contains some categories and sub-categories related to contents (kind of syllabus, organization and connection of units, and claims about the approach and methodology), layout (role of table of content and type of visual materials and their role) and additional materials (available kinds of additional materials). Micro evaluation contains unit grading (grading and sequencing of materials), reading comprehension skills (types of reading materials and strategies), other skills, lexis and grammar (approach to lexis and grammar), learners'/teachers' role (the role of teachers and learners, and allowance for differentiation), and user-friendliness (attract of the book, and auxiliary materials). The last stage in this framework is evaluating the compatibility of the book with the target curriculum/syllabus. The system of weighting in this framework is based on descriptions provided by the evaluators.

Rahimy's scheme is designed to evaluate reading comprehension textbooks and is piloted by the author. The main features of the checklist contain several detailed questions that focus the evaluators' attention toward what they should consider in their weightings. However, there is not clearly organized set of features or criterion to be followed while some items lack any more explanations or detailed questions such as: "other skills" (p.11), that really causes confusion. The scheme seems to provide a balance in its items while it considers several main and detailed features such as "how are the materials graded and sequenced" (p.11), and "is the book attractive" (p.12). The weighting system is based on provided descriptions by the users therefore it is totally subjective and might vary to a great extent from one to

other evaluator. Moreover, the comparison and graphic displays of the results will not be possible.

Discussion and Application

Through the recent decades a flourishing attempt has been made to develop textbook evaluation schemes and checklists. Yet the problem is that these results have not led to a wide use of the proposed schemes and checklists to carry out systematic and reliable evaluations. Although several reasons might be discernable in this regard, the fact is that mostly reasons other than the content and applicability of them are criticized, and they are rarely or never investigated for their strong and weak points.

This paper tried to explore several well-known checklists in terms of their content and specifically their practicality. The checklists were examined in terms of their format, scope, applied terms and concepts, weighting/rating systems, guidelines, and whether they were piloted by the developers.

With regard to the format, checklists should be easy to follow and user friendly while they should really avoid presenting cluttered and confusing materials. Moreover, the items, features, and factors should be signposted to help the users find their way through the material. This will also lead to more objective evaluations that prevents various interpretations of items and their features and/or factors.

Scope of the checklists should be broad enough to be adaptable or generalizable while including both valuable detailed features and broad concepts related to variety of aspects (physical aspects, content, methodology, etc). Anyhow, they should be flexible enough to be adapted in different contexts and for different aims.

The terms and concepts which are used in a checklist should be simple and clear enough and if not they should be accompanied by enough explanation or a set criteria to help the users understand what the author intended. This point has several advantages: it avoids confusion and subjectivity, leads to easier application of the items and avoids consuming more time finding criteria or understanding the concept, and impedes inattentive answers or skipping items. Moreover, the point should be

made that checklists should be designed in a way to be used by variety of stakeholders and not just experts.

Weighting/rating systems should be designed in a way that lead to more objective evaluations while they should be flexible enough to be used in variety of contexts and when adapting the checklists. The weighting/rating should be easy but based on criteria that make it reliable. The results should be clearly understandable and graphically displayable. Although the weighting/rating systems should avoid the use of specific terms, which cause variety of interpretations, they should be clear enough or might be defined in a way that all users have the same perception of them.

Guidelines are another important quality feature for checklists, which can be referred as a source for the users. They help the users find out how to use the checklist, how to weight/rate it, and how to adapt/change it. Also it helps the users to know which quality features and factors are intended by the author to be evaluated, and this will be a great help in decreasing the subjectivity of an evaluation. Piloting the checklist by the authors is another valuable point that helps them to judge and realize the weak points of their own work and eliminate it before it is widely published. Also this will help the checklist developers to examine their own checklists in terms of applicability and remove its probable obstacles.

The results of this review revealed that although the reviewed checklists have several strong points specifically regarding their format and scope, they mostly fail in terms of other features that lead to practicality. Furthermore, it is really worth considering that the available checklists are mostly designed to evaluate general English textbooks while they are not generalizable enough to be adapted to evaluate other English language textbooks. This study also focused on several specifically developed checklists for evaluating reading comprehension textbooks, while the results revealed that they also suffer from the same shortcomings. Therefore, based on the current study, the authors have developed a new checklist in which they have eliminated the mentioned defects. The checklist is a comprehensive reference specifically for evaluating reading comprehension textbooks while flexible enough to be used for other purposes too. For more details and the new checklist refer to Karamoozian and Riazi (2008).

References

Ansari, H., & Babaii, E. (2002). Universal characteristics of EFL/ESL textbooks: a step towards systematic textbook evaluation. *The Internet TESL Journal.* Vol. VIII, No. 2. Retrieved from: http://iteslj.org/Articles/Ansary-Textbooks/

Brown, H. D. (2001). *Teaching by principles: an interactive approach to language pedagogy.* (2nd ed.). New York, NY: Pearson Education Ltd.

Celce-Murcia, M. (2001). *Teaching English as a second or foreign language.* (3rd ed.). Boston, MA: Heinle & Heinle.

Garinger, D. (2001). Textbook evaluation. *TEFL Web Journal.* Retrieved from: http://www.teflweb-j.org/v1n1/garinger.html

Garinger, D. (2002). Textbook selection for the ESL classroom (Report No. EDO-FL-02-10). The U.S. Dep. of Education, Office of Educational Research and Improvement, National Library of Education (ERIC Document Reproduction in Service No. ED-99-CO-0008). Retrieved from: http://www.cal.org/ericcll/digest/0210garinger.html

Griffiths, C. (1995). Evaluating materials for teaching English to adult speakers of other languages. *Forum.* Vol.33, No.3. Retrieved from: http://exchanges.state.gov/forum/vols/vol33/no3/p50.htm

Harmer, J. (1998). *How to teach English: an introduction to the practice of English language teaching.* Harlow, UK: Longman

Hrehovcik, T. (2002). Foreign language textbook evaluation-methodological considerations. Retrieved from: http://www.univ.rzeszow.pl/fil_ang/usar1/sar_v1_18.pdf

Jahangard, A. (2007). The evaluation of the EFL materials taught at Iranian public high schools. *Karen's Linguistics Issues.* Retrieved from: http://www3.telus.net/linguisticsissues/bymonth.html

Karamoozian, F. M., & Riazi, A. (2008). Reading comprehension textbook evaluation checklist. Manuscript in progress for publication.

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

McGrath, I. (2002). Materials evaluation and design for language teaching. Edinburgh University Press.

Miekley, J. (2005). ESL textbook evaluation checklist. *The Reading Matrix.* Vol. 5, No.2. Retrieved from: http://www.readingmatrix.com/reading_projects/miekley/project.pdf

Peacock, M. (1997). Choosing the right book for your class. Retrieved from: http://www.sx.ac.uk/linguistics/pgr/egspll/volume1/PDFs/PEACOCK1.pdf

Rahimy, R. (2007). A coursebook evaluation. *ESP World.* Vol.6, Issue 2(15). Retrieved from: http://www.esp-world.info/contents.htm

Rivers, W. M. (1981). *Teaching foreign-language skills.* (2nd ed.). Chicago: The University of Chicago Press.

Sheldon, L. E. (1988). Evaluating ELT textbooks and materials. *ELT Journal, 42,* 237- 246.

Tomlinson, B. (1998). *Materials development in language teaching.* Cambridge: Cambridge University Press.

Ur, P. (1996*). A course in language teaching: practice and theory*. Cambridge: Cambridge University Press.

Williams, D. (1983). Developing criteria for textbook evaluation. *ELT Journal, 37,* 251- 255.

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