



Content list available at <http://ijltr.urmia.ac.ir>

*Iranian Journal  
of  
Language Teaching Research*



Urmia University

---

## Researcher and Non-researcher Teachers' Evaluation of ELT Materials: Converging or Diverging?

Brendalyn A. Manzano <sup>a,\*</sup>

<sup>a</sup> *Tarlac State University, Philippines*

---

### ABSTRACT

ELT materials evaluation is an important professional activity for ELT professionals. However, there are questions on who should perform the evaluation. Although there is a recognized importance of collaboration between researcher and non-researcher-teachers, some scholars have raised their concerns on the possible divergence between these two groups of teachers. Hence, this exploratory study sought to examine whether researcher-teachers (i.e., actively engaged in research) and non-researcher-teachers (i.e., not actively engaged in research) diverge or converge when evaluating ELT materials. The participants in this study involved five non-researcher-teachers and five researcher-teachers who evaluated the sample materials using a two-part questionnaire. These evaluators used an in-depth method because it is more focused and allows them to evaluate materials with depth. The scores assigned during validation were analyzed using descriptive statistics and inter-rater correlations. The findings revealed that the evaluation of the two groups of teachers was consistent in almost all evaluation criteria. However, non-researcher-teachers showed higher agreement among themselves than the researcher-teachers. These findings have implications for materials evaluation and classroom practices.

**Keywords:** ELT materials; language teaching; materials evaluation; materials development; textbooks

© Urmia University Press

---

### ARTICLE HISTORY

**Received:** 18 Nov. 2017

**Revised version received:** 27 Dec. 2017

**Accepted:** 10 June 2018

**Available online:** 1 July 2018

---

\* Corresponding author: Tarlac State University, Tarlac City, Philippines

Email address: [Brendalynmanzano@yahoo.com](mailto:Brendalynmanzano@yahoo.com)

© Urmia University Press

10.30466/ijltr.2018.120557

## Introduction

Nowadays, English language teaching (ELT) materials are becoming more and more sophisticated incorporating information and communication technology, role-plays, information-gap activities, authentic language samples, and realia (i.e., objects from real-life contexts and used for classroom instruction) that help make a connection between classrooms and real-life activities (Adel, Davoudi, & Ramezanzadeh, 2016; Anani Sarab, Monfared, & Safarzadeh, 2016; McDonough, Shaw & Masuhara, 2013; Nunan, 1999; Tomlinson, 2012; Tomlinson & Masuhara, 2017). Because of these developments, ELT materials continue to be a major tool for providing structure for teaching and learning (Garton & Graves, 2014; Hutchinson & Torres, 1994; Lee, 2015; McGrath, 2016). The acknowledged relevance of ELT materials necessitates materials developers to produce sound materials that incorporate the current principles of language teaching and learning. One crucial step toward ensuring this is through materials evaluation. The question, however, is whether researcher or non-researcher-teachers or both should perform the evaluation of ELT materials.

Although there is a recognized importance of collaboration between researcher and non-researcher-teachers, some scholars (e.g., Bell, 2007; Ellis, 1998a) have raised their concerns on the possible divergence between these two groups of ELT professionals. While non-researcher-teachers apply practical knowledge, researcher-teachers apply more technical knowledge (Ellis, 1998a). Technical knowledge refers to knowledge acquired from reflective inquiry and empirical investigation while practical knowledge relates to teachers' conceptualizations of language learning and teaching acquired from actual experience (Borg & Burns, 2008). This difference may result in conflicting beliefs between the two groups of teachers. For example, when non-researcher-teachers face classroom challenges, they would use practical knowledge over technical ones. With this, researcher-teachers and theorists tend to underestimate the non-researcher-teachers' ability to manage their own teaching as they undiscerningly follow methods they are trained in (Bell, 2007). Nonetheless, more experienced teachers would probably be able to integrate effectively these two types of knowledge (Barrot, 2016; Borg & Burns, 2008). It is in this context that this study was undertaken. Specifically, this paper sought to address the following questions: (1) To what extent do raters agree within each group? (2) To what extent do non-researcher-teachers and the researcher-teachers agree? To answer these questions, both quantitative and qualitative data were used.

## Review of the Literature

### *Materials Evaluation*

Materials evaluation is an important professional activity for ELT professionals (Garton & Graves, 2014; McDonough, Shaw & Masuhara, 2013) because it measures the relevance, appropriateness, and accuracy of the materials (Tomlinson, 2012). It is a complex process which considers pedagogical factors, cultural appropriateness, quality, quantity and types of exercise, methodology, skills, teachers' guide, variety, and pace (Chambers, 1997). Given the many factors to be considered during evaluation, ELT professionals need to have a comprehensive and deep understanding of language teaching and learning.

According to Ellis (1998b), materials evaluation can use various dimensions such as approach, purpose, focus, scope, evaluators, timing, and types of information. All of these dimensions apply to both macro (e.g., program) and micro evaluation (e.g., specific activities). Cunningsworth (1995) has proposed four criteria for evaluation textbooks: whether the textbook addresses

learners' needs, reflects the functional purpose of language that will be used by learners, facilitates learning processes and learners' needs without imposing a rigid method, and scaffolds learning.

Current practices reveal that materials evaluation mainly focuses on predictive evaluation (Ellis, 1997). This evaluation is usually carried out by experts or by teachers through the help of available guidelines and checklists which are organized parallel to how teachers go through the decision-making process. One limitation of such evaluation is its subjectivity due to the absence of clear-cut formula; hence, evaluation becomes more labor-intensive particularly when grading materials (Huang, 2007). As a reaction to this problem, Ellis (1997) proposed a micro-evaluation of materials in which a teacher chooses a specific task and subjects it to detailed empirical evaluation. A selected task can then be described in terms of its objectives, input, conditions, procedures, and outcomes.

Other methods have also been used as a guide in evaluating ELT materials. For instance, Littlejohn (2011) proposed a framework that focuses on examining the aspects of the materials as a pedagogic device and consists of two sections, publication and design. Publication refers to the physical features of the materials while design refers to the deeper underlying pedagogic principles, such as aims, principles of selection and sequencing, subject matter and its focus, teaching and learning activities, participation, learner and teacher roles, and roles of materials. Second, the framework suggests ways on how materials can be examined in three levels: what is there (level 1), what is required of users (level 2), and what is implied (level 3).

McDonough, Shaw, and Masuhara (2013) proposed that evaluation must begin with external evaluation; that is, examining the claims of the book, the introduction, and the table of contents. From the blurb and introduction of the book, evaluators would see who the intended users are and their proficiency level, the context of using materials, the presentation of language and teachable units, and the author's view of language teaching. Then, internal evaluation follows. It is done through an in-depth analysis of two or more units with respect to the presentation of skills, grading and sequencing of materials, types of texts, and relationship between tests and exercises.

Finally, McGrath (2016) proposed an evaluation using an impressionistic method, checklist method, and in-depth method. An impressionistic method deals with assigning global impression on the materials by reading the blurb, checking the table of contents, and looking at the topics, organizations, visuals, and layout and design. Unlike an impressionistic method, checklist method utilizes a checklist that contains items for verification and comparison. It is more systematic, cost-effective, follows a convenient format, and explicit; however, it lacks depth. This weakness of a checklist method is addressed by the in-depth method which is a more focused and specific form of evaluation. It uses representative samples (few lessons, chapters, or units only) for analysis using predetermined questions. However, this method is time-consuming and requires expertise. Nonetheless, the current study used this in-depth method because it is more focused and allows them to evaluate materials with depth.

### *Related Studies*

Although there have been several studies on materials evaluation (e.g., Arnold & Rixon, 2008; Chan, 2009; Dat, 2008; Kopperoinen, 2011; Masuhara & Tomlinson, 2008; Mol & Tin, 2008), there is an evident paucity of studies that compared the evaluation techniques of different groups of ELT teachers. One of these is the study of Johnson, Kim, Ya-Fang, Nava, Perkins, Smith, Soler-Canela, and Lu (2008) which investigated the textbook evaluation techniques of experienced and novice teachers. Three teachers participated in the study whose teaching experience ranged from 1 to 12 years. Using think-aloud protocol, the findings revealed that the three teachers used

different routes while they were evaluating the textbook. While the novice teacher used a back-and-forth approach, the more experienced teacher tended to be more focused, selective, and time efficient in his evaluation. The three teachers also differed on rating the usefulness of the textbook. The more experienced teacher provided wide-ranging issues while the other two teachers focused mainly on activities and flow of lesson.

A more recent study in this area was that of Tsagari and Sifakis (2014) who explored the views of teachers and book authors in the preparation of ELT course books. Using questionnaires and interview, their findings revealed that ELT materials production predominantly followed a top-down process. Moreover, teachers and authors differed in the way they perceived task variety, task demand, and task functions employed in the selected course books. For instance, the book authors believed that they provided sufficient variety of tasks in their course books to realize differentiation and that it is up to the authors to integrate such an approach in their respective classrooms. However, the teachers believed otherwise stating that course books should clearly specify how differentiation can be implemented using their course books.

Given the available literature, it appears that no study yet has compared the evaluation of ELT materials by researcher and non-researcher-teachers. Hence, this small-scale study intends to fill in this gap. It is hoped that the current study would extend our understanding of how researcher and non-researcher-teachers evaluate ELT materials. Given the exploratory nature of the current study, I do not aim to generalize findings but only to provide a framework for future large-scale studies on materials evaluation.

## Method

To address the research questions, the current study used a descriptive mixed method approach. This means that this paper focused on observing the measurable aspects of materials evaluation.

### *Participants*

The participants in this study involved five non-researcher-teachers (i.e., not active in research) and five researcher-teachers (i.e., active in research). The non-researcher-teachers are defined as those who have been teaching in Philippine ESL context for at least ten years in a university or basic education school, with experience in evaluating ELT materials or textbooks, specialized in language teaching, and have not been actively engaged in research for the past five years. Table 1 shows the teaching and professional background of the non-researcher-teachers.

Table 1  
*Non-researcher-teachers' Profile*

	N1	N2	N3	N4	N5
Specialization	Reading Education	Language Teaching	Reading Education	Language Education	Language Education
Years of teaching experience in ESL context	23	42	27	15	20
Experience in evaluating ELT materials	Yes	Yes	Yes	Yes	Yes

Researcher-teachers, on the other hand, are defined as those who have been teaching in ESL context for at least 10 years in a university, with experience in evaluating ELT materials or textbooks, specialized in language teaching, have been involved in research projects, and have published at least five research articles on applied linguistics and/or TESOL in the last five years. During the conduct of this study, they were active in both research and teaching. Note that both the research and non-researcher-teachers possessed at least a master's degree in applied linguistics, language teaching, or any related courses. Table 2 shows the profile of the five researcher-teachers.

Table 2  
*Researcher-teachers' Profile*

	R1	R2	R3	R4	R5
Specialization	Language acquisition; Discourse analysis	Language Testing and Assessment	Applied Linguistics	Reading and Literacy	Applied Linguistics
Years of teaching experience in ESL context	18	27	10	19	26
Experience in evaluating ELT materials	Yes	Yes	Yes	Yes	Yes
Number of articles published in refereed or abstracted journals	7	5	10	8	5
Article review experience	Yes	Yes	Yes	Yes	Yes

Table 3 summarizes the difference between the researcher-teachers and non-researcher-teachers. Since there are no studies that provide specific delineation between researcher and non-researcher-teachers, the author operationally specified the following criteria based on the hermeneutics.

Table 3  
*Difference between the Researcher-teachers and Non-researcher-teachers*

<b>Researcher-teachers</b>	<b>Non-researcher-teachers</b>
Have been involved in various research projects	No research activity for the past five years
With at least five research publications in reputable journals in applied linguistics and/or TESOL in the last five years	No research publications
With experience in reviewing research articles for journals	No experience in reviewing research articles for journals
With at least three years of experience as a panelist in thesis and dissertation defense	No experience as a panelist in thesis and dissertation defense
With at least three years of experience as a mentor for thesis and dissertation writing	No experience as a mentor for thesis and dissertation writing

### *Instrument*

The sample materials were subjected to evaluation using a two-part questionnaire (see Appendix A). The first part is a rating scale that determines whether the materials adhere to the current principles of language teaching and learning. The categories of the rating scale included the *content and pedagogy*, *coherence*, *applicability and usability*, and *clarity and appeal* (Barrot, 2015b). The rating scale used the following scores and verbal interpretations: 5 (to a very great extent), 4 (to a great extent), 3 (to a moderate extent), 2 (to a small extent), 1 (not at all), and NA (not applicable). The second part consists of open-ended questions on the weaknesses and strengths of the materials. The evaluators' responses to these questions provided the needed qualitative data to supplement the quantitative results. This instrument was validated by three experts who hold a doctoral degree in Applied Linguistics, with at least ten years of language teaching experience, and with experience in research and in preparing a similar instrument. These validators all agreed that the instrument covers all areas and addresses the specified research objectives.

To ensure consistent application of the questionnaire in evaluating the sample materials, each indicator was discussed to the evaluators through an orientation. In this way, the vague items and terms (e.g., differentiation, elements, constructivism) were clarified. The context in which the sample materials would be used, and their target learners were also discussed with the evaluators to determine whether the materials took into account learners' schema.

These teacher-made materials were reading and writing lessons for first-year college students of the English Communication Arts course. These materials were part of a thesis and validated by three experts. Each lesson is composed of the following elements: learning objectives, starter, self-assessment, discussion on reading (input 1), comprehension check, application, language focus (input 2), application, discussion on writing (input 3), individual and collaborative writing, reflection, and reinforcement activity. These materials were chosen because they reflect the current principles in language teaching and learning. Table 4 presents the sample materials that were evaluated.

Table 4  
*Sample Materials*

Lessons	Reading	Language	Writing
1	Identifying denotative and connotative meaning	Descriptive words	Static description
2	Sequencing events	Cohesive devices for chronology	Process description
3	Drawing conclusions	Cohesive devices for comparison and contrast	Comparison and contrast
4	Making inferences	Cohesive devices for cause and effect	Cause and effect
5	Building vocabulary through word analysis	Cohesive devices for exemplification	Definition
6	Distinguishing facts from opinions	Cohesive devices for conclusion	Argumentation

### *Procedure*

The evaluators used an in-depth method because it is more focused and allows them to evaluate materials with depth (McGrath, 2016). Each evaluator was given two representative chapters of ELT materials for college freshmen: (1) reading and writing argumentative essays and (2) reading and writing definition essays. Each chapter consists of three lessons: reading, grammar, and writing. Prior to evaluation, the evaluators were oriented on the evaluation procedure and the rating scale. The criteria and indicators in the rating scale were also explained to the evaluators to ensure that they have a similar understanding of its content. The evaluators were given three weeks to evaluate the materials and were recommended not to confer with one another so as not to influence the results of the evaluation. After the evaluators completed the evaluation, the accomplished rating scales were collected, and results were analyzed by the researcher.

### *Data Analysis*

The scores assigned during validation were subjected to descriptive statistics and inter-rater correlations. Specifically, the mean and standard deviation were used in quantitatively treating the scores assigned by the researcher and non-researcher-teachers. Fleiss's Kappa was also used to evaluate concordance between and among evaluators. It was used because the purpose of such treatment is not to measure reliability but agreement among multiple raters. The advantage of using Fleiss's Kappa is that it takes into account chance agreement (Barrot, 2015a; Stemler, 2004; Viera & Garrett, 2005). Kappa coefficient is interpreted using the following ranges: < 0 (poor agreement); 0.01–0.20 (slight agreement); 0.21–0.40 (fair agreement); 0.41–0.60 (moderate agreement); 0.61–0.80 (substantial agreement); 0.81–1.00 (almost perfect agreement). The level of acceptability was set at 0.41 (Yen, Kuppermann, Lillis, Monroe, Borgialli, Kerrey, Sokolove, Ellison, Cook, & Holmes, 2013).

The qualitative data provided by the evaluators were also analyzed in support of the quantitative data. The comments from the evaluators were coded and analyzed at the semantic level to determine the themes and categories. The comments that fall under the same theme (i.e., content and pedagogy, coherence, applicability and usability, and clarity and appeal) and of the same type (i.e., positive, negative, or neutral [Barkaoui, 2010]) were considered similar.

## **Results and Discussion**

The present study investigated whether there is an agreement within each group and between the two groups of materials evaluators in terms of evaluating the content and pedagogy, coherence, applicability and usability, and clarity and appearance of the sample materials. Both quantitative and qualitative data were provided to address these objectives.

Table 5  
*Evaluation Scores of Researcher and Non-researcher-teachers*

Indicators	Researcher-teachers' Scores		Non-researcher-teachers' Scores	
	Mean	SD	Mean	SD
<b>Content and Pedagogy</b>				
1	4.4	0.55	4.6	0.55
2	4.4	0.55	4.4	0.89
3	4.4	0.55	4.2	1.10
4	4.6	0.55	4.4	0.89
5	4.2	0.45	3.6	1.34
6	4.4	0.55	4.4	0.89
7	4.0	0.71	4.4	0.89
8	4.2	0.84	4.4	0.55
9	4.4	0.55	4.8	0.45
10	4.6	0.55	4.6	0.55
11	4.2	0.84	4.6	0.55
12	4.6	0.55	4.8	0.45
13	3.6	1.14	4.2	0.84
14	4.6	0.55	4.8	0.45
15	4.2	0.45	4.6	0.55
16	4.2	0.45	4.6	0.89
17	4.2	0.45	4.6	0.55
18	4.2	0.45	4.8	0.45
19	4.0	0.71	4.4	0.89
20	4.6	0.55	4.6	0.89
21	4.6	0.55	5.0	0.00
22	4.6	0.55	4.8	0.45
23	4.4	0.89	4.8	0.45
24	4.6	0.55	4.6	0.55
25	4.6	0.55	4.8	0.45
26	4.4	0.55	4.6	0.55
27	4.6	0.55	4.4	0.89
<b>Coherence</b>				
28	4.2	0.84	4.4	0.89
29	4.4	0.89	4.8	0.45
30	4.4	0.55	4.4	0.89
31	4.6	0.55	4.4	0.89

32	4.0	0.71	4.4	0.55
33	4.4	0.55	4.6	0.55
Applicability and Usability				
34	4.4	0.55	4.4	0.89
35	4.4	0.55	4.6	0.55
36	4.4	0.55	4.6	0.55
37	4.6	0.55	4.8	0.45
38	4.6	0.55	4.8	0.45
Clarity and Appearance				
39	4.6	0.55	4.4	0.89
40	4.6	0.55	4.6	0.55
41	4.4	0.89	4.8	0.45
42	4.4	0.55	4.2	0.84
43	4.6	0.55	4.8	0.45
44	4.6	0.55	4.6	0.55
Average	4.4	0.60	4.6	0.65

Table 5 presents the scores given by the evaluators to the sample materials. As can be seen, the non-researchers ( $\bar{x} = 4.60$ ) scored the materials higher than their researcher-teachers counterparts ( $\bar{x} = 4.40$ ). Although the scores given by these two groups of evaluators ranged from 3 to 5, most of these scores were either 4 or 5. Of the four categories of the criteria, *applicability and usability* received the highest aggregated score ( $\bar{x} = 4.56$ ), followed by *clarity and appeal* ( $\bar{x} = 4.55$ ), *content and pedagogy* ( $\bar{x} = 4.46$ ), and *coherence* ( $\bar{x} = 4.42$ )

One interesting finding that was obtained from this study is that items 5 (*The material reflects all of the components of communicative competence with due consideration on pragmatic competence*) and 13 (*The material takes into account the concept of differentiation*) posted the highest variation in scores among evaluators. Similar to coherence, the score variability may be explained by the fact that the pragmatic component of ELT materials is generally embedded and less observable (Eslami-Rasekh, 2005; Trosborg, 1995). As such, both the non-researcher and researcher-teachers had difficulties in identifying which part of the materials incorporated the development of pragmatic competence (i.e., how language is used in socioculturally appropriate ways) among learners and the extent to which it was incorporated into the materials. In the same way, differentiation seems to be sensitive to score variability due to difficulties in incorporating this concept into ELT materials even to experienced teachers and materials developers (Barrot, 2015b). The same is true when evaluating materials. Without seeing the actual use of the materials to their target learners, it would be challenging for the evaluators to gauge whether the materials faithfully adhere to the principle of differentiation. To compensate for this lack of information during evaluation, the evaluators tended to rate this aspect based on their own experiences which resulted in score variability.

Table 6  
*Inter-rater Agreement among Each Group of Evaluators*

Evaluators	Content & Pedagogy	Coherence	Applicability & Usability	Clarity and Appearance	Overall $\kappa$
Researchers	0.41	0.30	0.35	0.37	0.38
Non-researchers	0.45	0.38	0.46	0.45	0.44

Table 6 shows the interrater agreement among each group of evaluators. Findings reveal the non-researcher-teachers agree with one another ( $\kappa=0.44$ ) while the researcher-teachers did not ( $\kappa=0.38$ ). Among the four major criteria, both groups of evaluators were consistent in evaluating *Content and Pedagogy* ( $\kappa=0.41$ ;  $\kappa=0.45$ ). However, they failed to obtain an acceptable level of agreement in terms of coherence of the materials ( $\kappa=0.30$ ;  $\kappa=0.38$ ). This result may be explained by the nature of coherence which relates to the pragmatic features and culturally acceptable rhetorical organization, relationship among elements, structure, and sequence of materials. These features are naturally less observable compared to other elements of materials such as content, applicability and usability, and clarity and appearance. Assessing the coherence of materials also requires the evaluators to read recursively to check whether one section is logically linked to the immediately preceding and succeeding section and all other sections of the materials. It can also be posited that the differing specialization and theoretical orientation of the researcher-teachers may have contributed to the difference in the way they evaluated the materials. However, this claim needs to be confirmed in future studies. To address the difficulty in assessing the coherence of ELT materials, evaluators can use the scope and sequence which details how the elements of these materials are sequenced and linked with one another.

Table 7  
*Interrater Agreement between the Researcher and Non-researcher-teachers*

Content & Pedagogy	Coherence	Applicability & Usability	Clarity & Appearance	Overall $\kappa$
0.42	0.39	0.46	0.46	0.43

Table 7 reveals that the overall agreement between the non-researcher and researcher-teachers is acceptable but to a moderate extent only ( $\kappa=0.43$ ). In fact, these two groups of evaluators obtained an acceptable level of agreement in almost all areas except *coherence*. This result may be attributed to the similarities of their characteristics (i.e., teaching experience, materials evaluation experience, and specialization) except the research experience. The findings seem to support previous studies (Barrot, 2016; Borg & Burns, 2008) that experienced classroom teachers can effectively integrate theoretical and practical knowledge not only during classroom teaching but also during materials evaluation. One possible reason for this is that their extensive teaching experience guided them to generalize principled and evidence-based classroom practices needed for developing sound instructional materials. Finally, although their level of agreement was acceptable, it was not very high perhaps because of some intrinsic factors (e.g., embedded teaching framework, philosophies, and principles). As pointed out by Farrell and Vos (2018), teaching principles of teachers influence the way they view teaching.

The quantitative findings which show that agreement among non-researcher-teachers is higher than those of the researcher-teachers were supported by the comments provided by both groups. Generally, the similarities in the evaluators' comments were more frequent among the non-researcher-teachers. For example, three of them (N3, N4, and N5) mentioned that the activities in the materials were aligned to the learning objectives. N3 expressed that "the activities used in these materials were all targeting the set objectives" while N4 commented that "it's good to see that activities and learning objectives go together."

Others agreed that the materials incorporated the current principles in language teaching and learning (N1 and N2), were well thought off (N2 and N5), and used varied and ample activities (N2, N3, and N4). Although not part of the questionnaire, three non-researcher-teachers (N1, N2, and N3) also commented that the materials take into account Philippine issues. In fact, N2 articulated that “the provided materials expose students to current social issues.”

Aside from their positive feedback, the non-researcher-teachers also raised some opportunities for improvement. For instance, they pointed out that differentiation (N4 and N5) and emphasis on pragmatic competence (N3 and N4) were not sufficiently reflected in the materials. For instance, N4 commented that “the materials can further improve by making differentiation more concrete and easily applicable in a classroom setting” and “more activities targeting pragmatic competence be included by incorporating real-life discourse as texts.” Some of them (N1, N3, and N4) also commented that further discussion be added to the materials while others (N1 and N3) agreed that the starter needs to be simplified because they are too loaded.

Regarding the comments of the researcher-teachers, R2, R3, and R4 commented that the instructions for the writing tasks be revised for clarity and completeness. For instance, R2 noted that “some of the instructions are too complicated which makes it difficult for students to follow.” R3 and R4 further noted that there was a lack of tasks targeting the development of pragmatic competence. They also suggested that the discussion on writing argumentative essay be transferred immediately before the sample essays and that objectives on refutation be added. Some of them (R4 and R5) observed the lack of integrating differentiation into the lessons while others (R1, R4, and R5) commented that some concepts in the lessons were not sufficiently explained. As commented by R4, “these materials seem not to fully capture how differentiation can be realized, perhaps it’s because they are not as concrete as other pedagogical concepts.” R4 added that “more discussion should be added to provide students sufficient conceptual input.” R1 and R3 also agreed that the starter was too complicated. There were also few instances that the researcher-teachers provided positive comments on the materials. For instance, R2, R3, R4, and R5 all agreed that the materials provided ample activities that might help learners improve their writing skills. Interestingly, the researcher-teachers tended to be more critical than the non-researcher-teachers as reflected by the higher frequency of their negative comments.

## Conclusion

The present study sought to determine whether the researcher and non-researcher-teachers have diverging or converging perspectives during materials evaluation. The findings revealed that there was a moderate agreement between the two groups of evaluators. Within each group, the findings showed that the non-researcher-teachers tend to agree more with each other than the researcher-teachers. Findings also confirmed that the least agreement among evaluators occurred when evaluating the coherence of the materials. This was attributed to the fact that coherence is not directly observable and requires recursive reading among evaluators.

Several implications for materials evaluation and classroom practices can be deduced from this study. First, although it is desirable to have both the researchers and practitioners as materials evaluators, evaluation from non-researcher-teachers may be sufficient to ensure that ELT materials integrate the principles and theories in language teaching and learning because they can integrate both theoretical and practical knowledge during evaluation. Second, the findings revealed that there exists variability in attention that evaluators pay to different aspects of materials. There might also be a need to revisit the constructs for evaluating materials since some of them are sensitive to variability.

While the present study provided some interesting insights regarding materials evaluation, several limitations should be noted. First, only ten teachers participated in this study which makes the findings suggestive more than conclusive. Hence, future studies may include more participants of varying cultural background using multiple data collection methods such as interview, focus group discussion, and observation to obtain a clearer picture of the phenomena being investigated. Future researchers may also explore the factors that influence the evaluation processes and practices of both the researcher and non-researcher-teachers.

## References

- Adel, S. M. R., Davoudi, M., & Ramezanzadeh, A. (2016). A qualitative study of politeness strategies used by Iranian EFL learners in a class blog. *Iranian Journal of Language Teaching Research*, 4(1), 47–62.
- Anani Sarab, M. R., Monfared, A., & Safarzadeh, M. M. (2016). Secondary EFL school teachers' perceptions of CLT principles and practices: An exploratory survey. *Iranian Journal of Language Teaching Research*, 4(3), 109–130.
- Arnold, W., & Rixon, S. (2008). Materials for teaching English to young learners. In B. Tomlinson (Ed.), *English language learning materials* (pp. 38–58). London: Continuum International Publishing.
- Barkaoui, K. (2010). Do ESL essay raters' evaluation criteria change with experience? A mixed- methods, cross-sectional study. *TESOL Quarterly*, 44(1), 31–57. doi: 10.5054/tq.2010.214047
- Barrot, J. S. (2015a). Self- and teacher assessment of speaking performances: An investigation of interrater consistency and agreement. *Journal of Language and Literature*, 6(4), 193–200.
- Barrot, J. S. (2015b). A sociocognitive-transformative instructional materials design model for second language (L2) pedagogy in the Asia Pacific: Development and validation. *The Asia-Pacific Education Researcher*, 24(2), 283–297.
- Barrot, J. S. (2016). Examining the teaching beliefs and practices of experienced ESL teachers: A sociocognitive-transformative perspective. *3L: The Southeast Asian Journal of English Language Studies*, 22(1), 153–163.
- Bell, D. (2007). Do teachers think that methods are dead? *ELT Journal*, 61, 135–143. doi: 10.1093/elt/ccm006
- Borg, S., & Burns, A. (2008). Integrating grammar in adult TESOL classrooms. *Applied Linguistics*, 29(3), 456–482. doi: 10.1093/applin/amn020
- Chambers, F. (1997). Seeking consensus in coursebook evaluation. *ELT Journal*, 51(1), 29– 35. doi: 10.1093/elt/51.1.29
- Chan, C. (2009). Forging a link between research and pedagogy: A holistic framework for evaluating business English materials. *English for Specific Purposes*, 28(2), 125–136. doi: 10.1016/j.esp.2008.12.001

- Cunningsworth, A. (1995). *Choosing your coursebook*. Oxford: Pergamon.
- Dat, B. (2008). ELT materials used in Southeast Asia. In B. Tomlinson (Ed.), *English language learning materials* (pp. 263–280). London: Continuum International Publishing.
- Ellis, R. (1997). The empirical evaluation of language teaching materials. *ELT Journal*, 51(1), 36–42. doi: 10.1093/elt/51.1.36
- Ellis, R. (1998a). Teaching and research: Options in grammar teaching. *TESOL Quarterly*, 32(1), 39–60. doi: 10.2307/3587901
- Ellis, R. (1998b). The evaluation of communicative tasks. In B. Tomlinson (Ed.), *Materials Development in Language Teaching* (pp. 217–238). Cambridge: Cambridge University Press.
- Eslami-Rasekh, Z. (2005). Raising the pragmatic awareness of language learners. *ELT Journal*, 59(3), 199–208. doi: 10.1093/elt/cci039
- Farrell, T. S., & Vos, R. (2018). Exploring the principles and practices of one teacher of L2 speaking: The importance of reflecting on practice. *Iranian Journal of Language Teaching Research*, 6(1), 1–15.
- Garton, S., & Graves, K. (2014). Identifying a research agenda for language teaching materials. *The Modern Language Journal*, 98(2), 654–657.
- Huang, H. (2007). Vocabulary learning in an automated graded reading program. *Language Learning & Technology*, 11(3), 64–82.
- Hutchinson, T., & Torres, E. (1994). The textbook as agent of change. *ELT Journal*, 48(4), 315–328. doi: 10.1093/elt/48.4.315
- Johnson, K., Kim, M., Ya-Fang, L., Nava, A., Perkins, D., Smith, A. M., Soler-Canela, O., & Lu, W. (2008). A step forward: investigating expertise in materials evaluation. *ELT Journal*, 62(2), 157–163. doi: 10.1093/elt/ccl021
- Kopperoinen, A. (2011). Accents of English as a lingua franca: a study of Finnish textbooks. *International Journal of Applied Linguistics*, 21(1), 71–93. 10.1111/j.1473-4192.2010.00263.x
- Lee, B. (2015). EFL learners' perspectives on ELT materials evaluation relative to learning styles. *RELC Journal*, 46(2), 147–163.
- Littlejohn, A. (2011). The analysis of language teaching materials: Inside the Trojan Horse. In B. Tomlinson (Ed.), *Materials Development in Language Teaching* (pp. 190–216). Cambridge: Cambridge University Press.
- Masuhara, H., & Tomlinson, B. (2008). Materials for general English. In B. Tomlinson (Ed.), *English Language Learning Materials* (pp. 17–37). London: Continuum International Publishing.
- McDonough, J., Shaw, C., & Masuhara, H. (2013). *Materials and methods in ELT: A teacher's guide* (3rd ed.). West Sussex: John Wiley & Sons.

- McGrath, I. (2016). *Materials evaluation and design for language teaching* (2nd ed.). Edinburg: Edinburg University Press.
- Mol, H., & Tin, T. (2008). EAP materials in New Zealand and Australia. In B. Tomlinson (Ed.), *English language learning materials* (pp. 74–99). London: Continuum International Publishing.
- Nunan, D. (1999). *Second language teaching and learning*. Boston, MA: Heinle and Heinle.
- Stemler, S. (2004). A comparison of consensus, consistency, and measurement approaches to estimating interrater reliability. *Practical Assessment, Research and Evaluation*, 9(4), 1–19. Retrieved March 23, 2016 from <http://PAREonline.net/getvn.asp?v=9andn=4>
- Tomlinson, B. (2012). Materials development for language learning and teaching. *Language Teaching*, 45(2), 143–179. doi: 10.1017/S0261444811000528
- Tomlinson, B., & Masuhara, H. (2017). *The complete guide to the theory and practice of materials development for language learning*. New York: John Wiley & Sons.
- Trosborg, A. (1995). *Interlanguage pragmatics*. New York: Mouton de Gruyter.
- Tsagari, D., & Sifakis, N. C. (2014). EFL course book evaluation in Greek primary schools: Views from teachers and authors. *System*, 45, 211–226.
- Viera, A., & Garrett, J. (2005). Understanding interobserver agreement: The kappa statistic. *Fam Med*, 37(5), 360–363.
- Yen, K., Kuppermann, N., Lillis K., Monroe, D., Borgialli, D., Kerrey, B.T., Sokolove, P., Ellison, A., Cook, L., & Holmes, J. (2013). Interobserver agreement in the clinical assessment of children with blunt abdominal trauma. *Academic Emergency Medicine*, 20(5), 426–432. doi: 10.1111/acem.12132

### Acknowledgment

This research project was fully funded (Grant No. AEIRS2017061R-01) by the Association of Educators for Interrelated Research Studies, a non-government organization accredited as a continuing professional development provider for teachers by the Philippine Regulations Commission. The author's heartfelt thank is also extended to Tarlac State University for providing incentive for this paper. Sincere thanks are also given to the participants who shared their precious time to provide the substantial data. Finally, the author is thankful to the anonymous reviewers whose constructive comments and insightful suggestions helped improve this paper and to the editors of IJLTR for publishing this paper.

**Brendalyn A. Manzano** is an associate professor and chairperson of the English Department of Tarlac State University. She holds a PhD in Linguistics with specialization in Applied Linguistics. Her research interests include second language acquisition, translation, language pedagogy, and ELT materials evaluation.

## Appendix A

Part I.		To a Very Great Extent	To a Great Extent	To a Moderate Extent	To a Small Extent	Not at All	Not Applicable
<b>CONTENT AND PEDAGOGY</b>							
1.	The material takes into account the concept of constructivism (i.e., allows learners to make meaning).	5	4	3	2	1	NA
2.	The material takes into account both the social and cognitive factors in language learning and development.	5	4	3	2	1	NA
3.	The material aims to develop learners' language competence necessary for social and self-transformation.	5	4	3	2	1	NA
4.	The material reflects the development and integration of multiple literacies.	5	4	3	2	1	NA
5.	The material reflects all of the components of communicative competence with due consideration on pragmatic competence.	5	4	3	2	1	NA
6.	The material reflects social interaction and collaboration for language learning.	5	4	3	2	1	NA
7.	The material takes into account the concept of spiral progression	5	4	3	2	1	NA
8.	The material incorporates the use of information and communication technology with emphasis on integrative CALL.	5	4	3	2	1	NA
9.	The material takes into account the concept of contextualization.	5	4	3	2	1	NA
10.	The material promotes authentic language experience among learners.	5	4	3	2	1	NA
11.	The material takes into account the learners' schema in the teaching-learning process.	5	4	3	2	1	NA
12.	The material incorporates the concept of reflective learning and critical thinking.	5	4	3	2	1	NA
13.	The material takes into account the concept of differentiation.	5	4	3	2	1	NA
14.	The model emphasizes both the process and the product.	5	4	3	2	1	NA
15.	The material incorporates the integration of teaching grammar and vocabulary to the teaching of macroskills in an authentic context in both explicit instruction and incidental learning.	5	4	3	2	1	NA
16.	The material takes into account the integration of all macroskills: speaking, listening, reading, writing, viewing, and representing.	5	4	3	2	1	NA
17.	The material takes into account form-meaning connections in language learning.	5	4	3	2	1	NA
18.	The material reflects the integration of a form-focus instruction and meaning-based activities.	5	4	3	2	1	NA
19.	The material takes into account the learners' social realities and psycholinguistic abilities in selecting teaching and learning content and activities.	5	4	3	2	1	NA
20.	The material incorporates various forms of assessment (self-, peer, and teacher) in both traditional and alternative methods.	5	4	3	2	1	NA
21.	The learning objectives are clearly defined.	5	4	3	2	1	NA
22.	The learning objectives are aligned to the goal of the curriculum.	5	4	3	2	1	NA
23.	The material takes into account the various forms of input.	5	4	3	2	1	NA
24.	The material takes into account the various forms of tasks.	5	4	3	2	1	NA
25.	The material links tasks and assessment.	5	4	3	2	1	NA
26.	The material is free from stereotypes.	5	4	3	2	1	NA
27.	All major elements (e.g., learning objectives, preparatory stage, input, main tasks, and reinforcement activities) are included in the materials.	5	4	3	2	1	NA
<b>COHERENCE</b>							
28.	The elements are properly sequenced.	5	4	3	2	1	NA
29.	Relationships among elements are clear.	5	4	3	2	1	NA
30.	All elements are properly linked to one another.	5	4	3	2	1	NA

31.	The elements are clearly and logically grouped and sub-grouped.	5	4	3	2	1	NA
32.	The organization is simple enough to be understood.	5	4	3	2	1	NA
33.	The material allows a recursive (non-linear) process of learning.	5	4	3	2	1	NA
<b>APPLICABILITY AND USABILITY</b>							
34.	The elements of the material can be adopted for use in other lessons/topics.	5	4	3	2	1	NA
35.	The material allows flexibility and creativity on the part of the teacher.	5	4	3	2	1	NA
36.	The material does not conflict with any established teaching frameworks.	5	4	3	2	1	NA
37.	The material can accommodate acceptable and established principles and theories in language teaching and learning.	5	4	3	2	1	NA
38.	The material operates in an ESL context.	5	4	3	2	1	NA
<b>CLARITY AND APPEARANCE</b>							
39.	The language is appropriate to its target users.	5	4	3	2	1	NA
40.	The language is simple enough for easy comprehension and interpretation.	5	4	3	2	1	NA
41.	The illustrations/graphs are appropriately used.	5	4	3	2	1	NA
42.	No elements and sub-elements are overlapping and redundant.	5	4	3	2	1	NA
43.	All labels/headings are clear and comprehensible.	5	4	3	2	1	NA
44.	Overall appearance is pleasing.	5	4	3	2	1	NA
Part II.							
What are the strengths of the materials?							
_____							
What are the weaknesses of the materials?							
_____							
_____							