Multi-Criteria Evaluation and Benchmarking for Young Learners' English Language Mobile Applications in Terms of LSRW Skills

N. K. IBRAHIM[1], HAMSA HAMMED[2], A. A. ZAIDAN[1], B. B. ZAIDAN[1], O. S. ALBAHRI[1], M. A. ALSALEM[3], R. T. MOHAMMED[4,5], ALI NAJM JASIM[1], ALI H. SHAREEF[1], N. S. JALOOD[1], M. J. BAQER[1], SHAHAD NIDHAL[6], E. M. ALMAHDI[1], AND MUSAAB ALAA[1]

[1]Department of Computing, Universiti Pendidikan Sultan Idris, Tanjong Malim 35900, Malaysia
[2]Faculty of Human Development, Sultan Idris Education University, Tanjung Malim 35900, Malaysia
[3]Department of Management Information System, College of Administration and Economic, University of Mosul, Mosul 41001, Iraq
[4]Department of Computer Science, Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, Seri Kembangan 43400, Malaysia
[5]Geomatika University College, Kuala Lumpur 54200, Malaysia
[6]Department of Computer Technology Engineering, Dijlah University, Baghdad 10022, Iraq

Corresponding author: Hamsa Hammed (hamsahameed@fpm.upsi.edu.my)

ABSTRACT This study proposes an evaluation and benchmarking decision matrix (DM) on the basis of multi-criteria decision making (MCDM) for young learners' English mobile applications (E-apps) in terms of listening, speaking, reading and writing (LSRW) skills. Benchmarking E-apps for young learners is challenging due to (a) multiple criteria, (b) criteria importance and (c) data variation. The DM was constructed on the basis of the intersection amongst evaluation criteria in terms of LSRW and E-apps for young learners. The criteria were adopted from a preschool education curriculum standard. The DM data included six E-apps as alternatives and 17 skills as criteria. Thereafter, the six E-apps were evaluated by distributing a checklist form amongst six English learning experts. These apps were subsequently benchmarked by utilising MCDM methods, namely, best–worst method (BWM) and technique for order of preference by similarity to ideal solution (TOPSIS). BWM was used for criterion weighting, whereas TOPSIS was employed to benchmark and rank the apps. TOPSIS was utilised in two contexts, namely, individual and group. In the group context, internal and external aggregations are applied. Mean was computed to ensure that the E-apps undergo a systematic ranking for objective validation. This study provides scenarios and a benchmarking checklist to evaluate and compare the proposed work with six relative studies. Results indicated that (1) BWM is suitable for criteria weighting. (2) TOPSIS is suitable for benchmarking and ranking E-apps. Moreover, the internal and external TOPSIS group decision making exhibited similar findings, with the best app being 'Montessori' and the worst app being 'FunWithFlupe.' (3) For objective validation, remarkable differences were observed amongst the group scores, which indicate that the internal and external ranking results are identical. (4) In the evaluation, the proposed DM revealed advantages over the six relative studies by 40.00%, 53.33%, 40.00%, 46.67%, 46.67% and 46.67%.

INDEX TERMS Language learning app evaluation, language learning app assessment, language teaching/learning strategies.

I. INTRODUCTION

Currently, the world is considered a global village where individuals communicate with one another through a popular language, which is known as English, a common language in the international community [1]. English is utilised at a useful level by approximately 1.75 billion people worldwide, which indicates that one out of four people speaks this language [2]. By 2020, two billion people are predicted to speak English or learn to use it.

The associate editor coordinating the review of this manuscript and approving it for publication was Xiao-Sheng Si.

N. K. Ibrahim *et al.*: Multi-Criteria Evaluation and Benchmarking for Young Learners' English Language Mobile Applications

IEEE *Access*

The early childhood stage is considered the most rapid period of development in human life that is important for the holistic and healthy cognitive, emotional and physical growth of children [3]. English education in early childhood can facilitate effective English learning in young children and support formal English learning in primary school. Thus, inspiring the learning interest of children is critical and represents the key in improving teaching quality [4].

Technological development with mobile devices and application adoption have translated into several opportunities for children to learn English [5]. Mobile devices can promote motivation in children [6], make the process of language learning fun and enjoyable and help learners positively develop language skills [5], [7]. Mobile devices have provided a unique learning environment and a vast opportunity for young English language learners to practice and learn English [8]. Extensive findings reveal a total development towards utilising mobile phones to promote the efficiency and eminence of mobile learning for young learners. Therefore, mobile phones are becoming popular for their role as new and effective learning tools [9], [10]. Recently, mobile applications (apps) are constantly increasing and considered communal tools for learning [11]. They are the latest technological developments that aid English learning [12] and considered one of the preferred environments for children to learn and practice a language [13]. Although many apps are available in stores, not all of them are widely used [11]. A great number of available English learning apps are provided for children, [6] but whether all of these apps are designed with a theoretical approach cannot be concluded [14]. No quality control is employed to assess their content [15], resulting in users' difficulty in choosing the right app [16]. Therefore, evaluating and benchmarking these apps is necessary when selecting the best one. Benchmarking refers to a standard or a group of standards utilised as a point of reference for evaluating the performance or level of quality when compared against others [17], [18].

In our context, English learning apps' benchmarking process is considered a challenge because each English learning app content must be evaluated in many aspects. Each aspect also includes a set of criteria that should be considered. These aspects are embodied by three main skills adopted from the 2016 KSPK standard as follows:

(1) Listening and Speaking (stimulus given, rimes, poems and rhymes, stories, favourite things and activities, oral texts, familiar activities and experiences, stories heard, daily situations) criteria;
(2) Reading (Alphabet letters, simple phrases, simple sentences, texts) criteria; and
(3) Writing (copy legible phrases, copy legible sentences, ideas and information communication and legible writing) criteria.

Benchmarking for English learning apps has been facing issues due to several important aspects. Such issues are multiple criteria, criterion importance and data variation.

Firstly, the multiple criteria issue involves many criteria that affect the benchmarking process of English learning apps. This study involves three main aspects with 17 criteria. Therefore, all these criteria should be simultaneously considered [19]–[21]. Secondly, criterion importance depicts that one criterion may be preferred more than the others [22]–[24]. That is, these criteria are not at the same level of importance [25]–[27]. Thirdly, data variation term means that various alternatives can be represented as a set of various data during the benchmarking process. Specifically, each single English learning app can be assigned with a single value for each criterion [28]–[30]. Data variation amongst criteria causes a problem with which decision makers cannot compare an app with other apps [31]–[33], [35]. This appointed area of evaluation and benchmarking is considered a multiple criteria problem. Thus, this research provides a new decision-making solution to evaluate and benchmark English learning apps on the basis of the three different aspects of listening, speaking, reading and writing (LSRW) by using multi-criteria decision making (MCDM). This study can help English language teachers and designers to understand how contents of English courses should be presented. Moreover, this study can assist parents and kindergarten teachers for screening and selecting suitable and reliable English learning apps. Figure 1 illustrates the framework of the paper. The remainder of this paper is structured as follows. Section II discusses the literature review. Section III describes the decision-making methodology for evaluating and benchmarking English learning apps. Section IV presents the results and discussion. Section V discusses the validation and evaluation processes. Finally, Section VI provides the conclusion.

II. LITERATURE REVIEW

In this study, a representative and comprehensive systematic review of 618 journal articles about English learning mobile apps evaluation is conducted to determine the research gaps. Our critical review only suggests six studies related to evaluating apps for English learning. This section discusses and reviews these papers to identify the criteria and approaches used for evaluation. Amongst these six papers, five (5/6) are related to adults, rather than early childhood. In these five papers, the authors evaluated the apps on the basis of several criteria. For instance, the authors in [36] and [37] assessed English mobile applications (E-apps) according to ease of use, functionality, design/layout, usability and usefulness amongst others. The authors in these two papers (2/5) completely ignored content evaluation. In the last three papers [9], [14], [38], the apps were evaluated using almost the same criteria and content with other related studies included in this paper. With regard to content evaluation, the authors in [9] designed an app to promote idioms and vocabularies, which focus on whether the content is nicely presented, visibly described to users, noteworthy and helps learners to become further engaged and motivated.
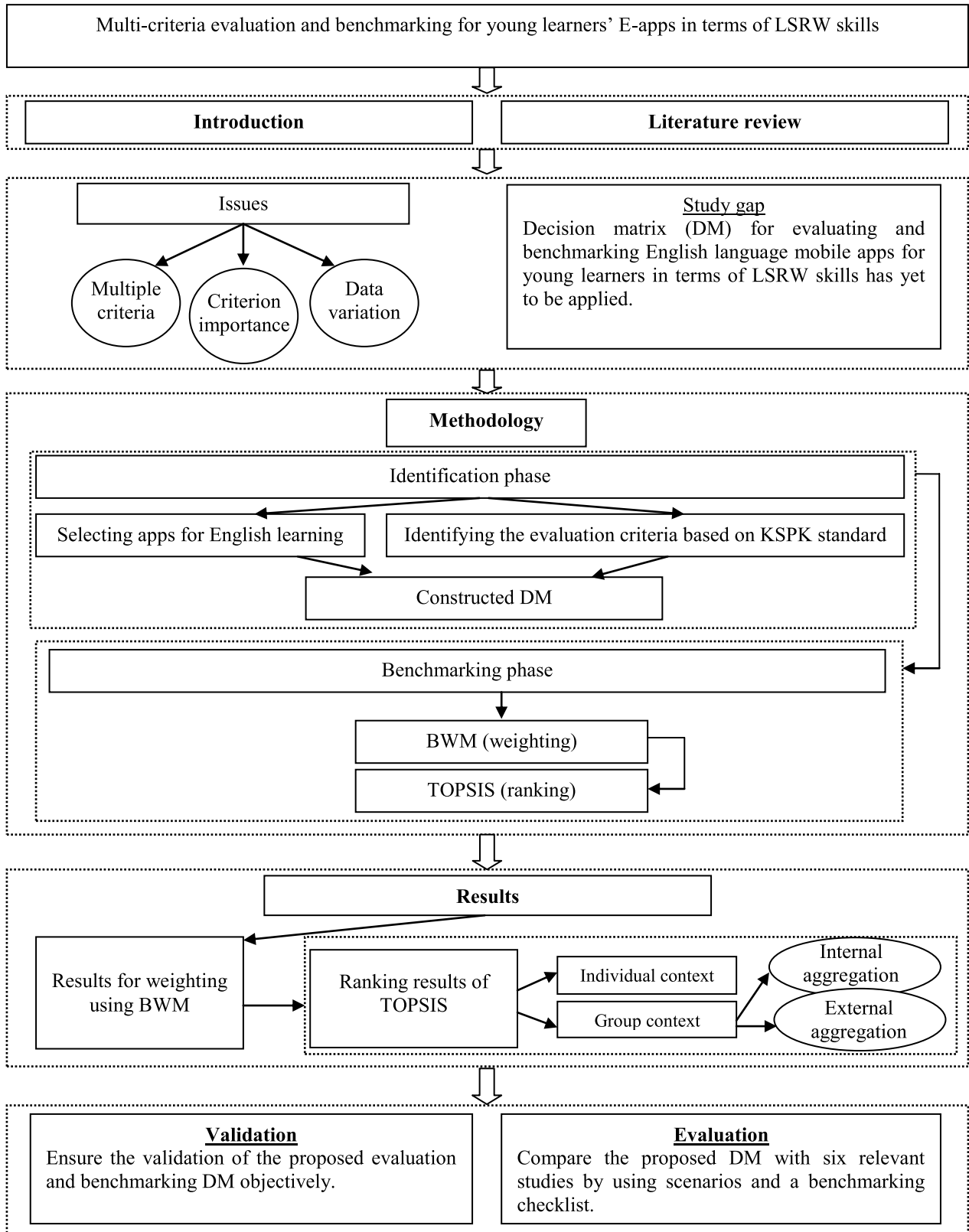
IEEE *Access*

N. K. Ibrahim *et al.*: Multi-Criteria Evaluation and Benchmarking for Young Learners' English Language Mobile Applications

Multi-criteria evaluation and benchmarking for young learners' E-apps in terms of LSRW skills

Introduction   **Literature review**

Issues

Multiple criteria   Criterion importance   Data variation

Study gap
Decision matrix (DM) for evaluating and benchmarking English language mobile apps for young learners in terms of LSRW skills has yet to be applied.

Methodology

Identification phase

Selecting apps for English learning   Identifying the evaluation criteria based on KSPK standard

Constructed DM

Benchmarking phase

BWM (weighting)

TOPSIS (ranking)

Results

Results for weighting using BWM   Ranking results of TOPSIS   Individual context   Group context   Internal aggregation   External aggregation

Validation
Ensure the validation of the proposed evaluation and benchmarking DM objectively.

Evaluation
Compare the proposed DM with six relevant studies by using scenarios and a benchmarking checklist.

FIGURE 1. Framework of the study.

N. K. Ibrahim *et al.*: Multi-Criteria Evaluation and Benchmarking for Young Learners' English Language Mobile Applications

IEEE*Access*

In [14], content quality evaluation focused on the content presentation, whereas [38] merely emphasised the pronunciation skill, which is a part of speaking [39]. Moreover, evaluation for the three other skills was not conducted. Only one (1/6) paper was related to early childhood (kindergarten students). In [40], an app was evaluated for its effectiveness by asking participants about the user interface design, materials and functionality of the app. Although the prototype was evaluated through learning materials, vocabulary was the focus, rather than the four main language LSRW skills. The cited critical review revealed that no study has been conducted to identify and construct a DM. This matrix can be used to evaluate and benchmark young learners' mobile apps for learning English as a second language in terms of LSRW skills as an integrated platform on the basis of the KSPK standard.

III. METHODOLOGY

A two-phase methodology of benchmarking the English learning apps is employed. The first phase (identification) intends to construct the DM on the basis of the intersection between criteria and English learning apps. The second phase (benchmarking) includes English learning app benchmarking and ranking based on the integrated Best–Worst Method (BWM)–Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) methods.

A. IDENTIFICATION PHASE

This phase aims to identify DM on the basis of the intersection between multi-evaluation criteria in terms of LSRW skills and young learners' E-apps. In any MCDM cases, existing substantial terms must be defined, such as the alternatives, the criteria and the decision or evaluation matrix [41]–[45].

1) SELECT MOBILE APPS FOR ENGLISH LEARNING (ALTERNATIVES)

Alternatives refer to the various (usually finite) choices of decision makers for the considered problem [46]–[49]. We have selected six available apps for early childhood English learning as alternatives for this study. These apps are suitable for students aged 5 and above and are based on the access dates in September 2018. Specifically, these apps are: Lingokids, Fun English, FunWithFlupe, First Words, Montessori and Spelling Bee. This list of apps is not comprehensive of all available early childhood English learning apps but is a representative group of common apps in literature together with open source repositories.

2) IDENTIFY THE EVALUATION CRITERIA ON THE BASIS OF LSRW SKILLS

Criteria refer to the different dimensions through which the alternatives can be viewed [50], [51]. The criteria used in this study are identified from the 2016 KSPK standard, which includes main and sub-criteria. Figure 2 illustrates the criteria employed in this study and represents the final set with
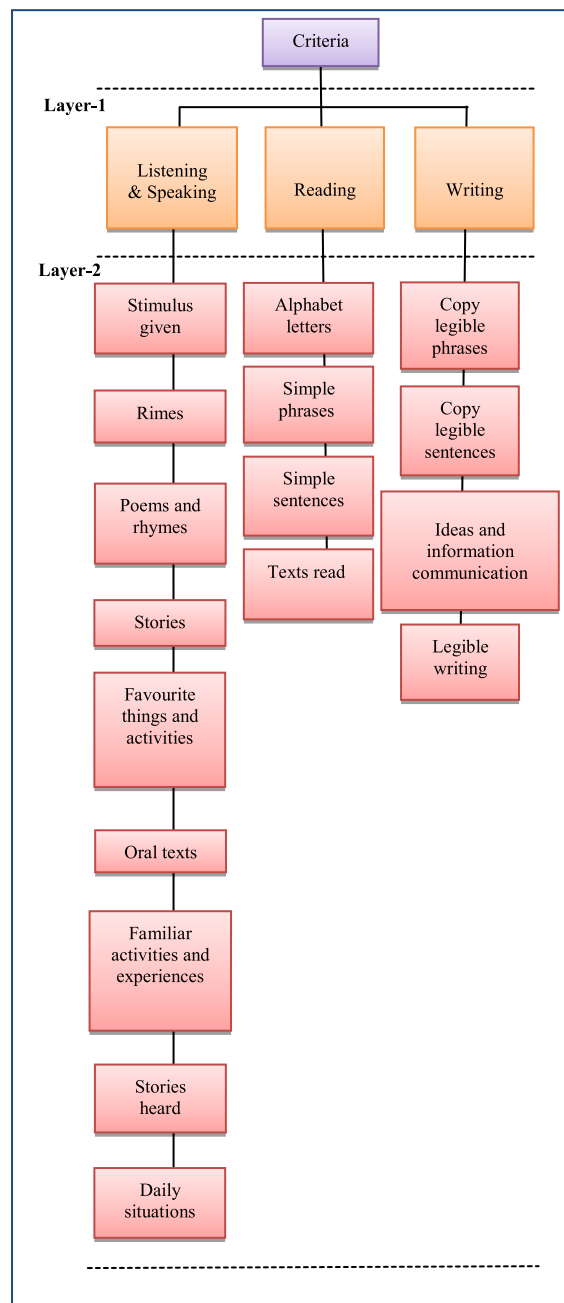

FIGURE 2. Levels of evaluation criteria for evaluating the English learning apps based on the 2016 KSPK standard.

respect to experts' views, including the four main language LSRW skills.

Listening and Speaking skills are combined as one aspect in the 2016 KSPK standard and is therefore evaluated through nine criteria, namely, stimulus given, rimes, poems and rhymes, stories, favourite things and activities, oral texts, familiar activities and experiences, stories heard and daily situations.

The reading aspect in the 2016 KSPK standard consists of eight criteria, namely, alphabet letters, sounds in a word, blend sounds, frequency/sight words, simple phrases, simple

IEEE Access

N. K. Ibrahim *et al.*: Multi-Criteria Evaluation and Benchmarking for Young Learners' English Language Mobile Applications

TABLE 1. Proposed DM.

Main Criteria	Listening and Speaking									Reading				Writing			
Sub-criteria (apps)	Stimulus given (sg)	Rimes (ri)	Poems and rhymes (p&r)	Stories (st)	Favourite things and activities (ft&a)	Oral texts (ot)	Familiar activities and experiences (fa&e)	Stories heard (sh)	Daily situations (ds)	Alphabet letters (al)	Simple phrases (sp)	Simple sentences (ss)	Text reading (tr)	Copy legible phrases (clp)	Copy legible sentences (cls)	Idea and information communication (i&ic)	Legible writing (lw)
App1 (A1)	A1 (sg)	A1 (ri)	A1 (p&r)	A1 (st)	A1 (ft&a)	A1 (ot)	A1 (fa&e)	A1 (sh)	A1 (ds)	A1 (al)	A1 (sp)	A1 (ss)	A1 (tr)	A1 (lpc)	A1 (lsc)	A1 (i&ic)	A1 (lw)
App2 (A2)	A2 (sg)	A2 (ri)	A2 (p&r)	A2 (st)	A2 (ft&a)	A2 (ot)	A2 (fa&e)	A2 (sh)	A2 (ds)	A2 (al)	A2 (sp)	A2 (ss)	A2 (tr)	A2 (lpc)	A2 (lsc)	A2 (i&ic)	A2 (lw)
App3 (A3)	A3 (sg)	A3 (ri)	A3 (p&r)	A3 (st)	A3 (ft&a)	A3 (ot)	A3 (fa&e)	A3 (sh)	A3 (ds)	A3 (al)	A3 (sp)	A3 (ss)	A3 (tr)	A3 (lpc)	A3 (lsc)	A3 (i&ic)	A3 (lw)
—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
App n (An)	An (sg)	An (ri)	An (p&r)	An (st)	An (ft&a)	An (ot)	An (fa&e)	An (sh)	An (ds)	An (al)	An (sp)	An (ss)	An (tr)	An (lpc)	An (lsc)	An (i&ic)	An (lw)

sentences, independency and texts read. Experts view four of these criteria, namely, sounds in a word, blend sounds, frequency/sight words and independency, as unsuitable for early childhood English learning apps. Thus, these four criteria are disregarded in this study. The final criteria set under the reading aspect consists of the four remaining criteria, which are alphabet letters, simple phrases, simple sentences and texts read. Moreover, four criteria are utilised to evaluate the writing aspect, namely, copy legible phrases, copy legible sentences, ideas and information communication and legible writing.

3) CONSTRUCT THE DM

In this section, an intersection is designed between mobile apps (alternatives) and identified aspects, which are LSRW skills and their criteria. Table 1 presents the DM. Data presentation is needed in this stage to fulfil the proposed DM. Practically, each app comprises many subskills that must be considered. The three specific issues mentioned in the Introduction section are addressed as follows: (1) multiple criteria indicate that apps are evaluated with respect to the 17 subskills; (2) data variation refers to the different values of the apps' subskills, but this variation of values results in

ranking and selection difficulty; and (3) the importance of criteria, which suggesting that varied weights exist between main skills and subskills.

4) EVALUATE THE DM

The six English learning apps are evaluated by a panel of experts according to the identified LSRW skill criteria. The next subsections describe experts' selection, checklist form for evaluation and evaluation procedures undertaken for English learning mobile apps based on LSRW skill criteria.

a: EXPERT SELECTION

The panel of experts consists of six English learning lecturers in early childhood learning department/Faculty of Education and Human Development in Universiti Pendidikan Sultan Idris. All these experts have been teaching early childhood English courses for over five years, as characterised in studies [1], [14], [52] and [53]. Thus, they have had rich experience in teaching English. Checklist forms are distributed amongst them. The experts participate in three stages. Firstly, they evaluate and test the format and the content of the checklist form. Secondly, as respondents of this study, they help us in gathering data for analysis by answering the checklist

N. K. Ibrahim et al.: Multi-Criteria Evaluation and Benchmarking for Young Learners' English Language Mobile Applications

IEEE Access

TABLE 2. Weaknesses and strengths of common MCDM methods.

Method	Strengths	Weaknesses
HAW & WSM	▪ easy to understand and apply	▪ criteria weights are assigned arbitrarily ▪ difficult to adopt in the case of many criteria ▪ uses common numerical scaling to calculate the final score
WPM & MEW	▪ their ability to eliminate any item is measured ▪ the use of relative values, rather than actual ones	▪ these two methods do not offer any solution with an equal DM weight
ANP	▪ produces a full understanding of the significance level that a criterion can take regarding its correlation with other criteria ▪ allows for measurement of judgements' consistency, which is impossible to evaluate in the method that assigns weights by compromise ▪ helps assign weights by breaking down the problem into smaller parts, so that a group of experts can have a manageable discussion because only two criteria are compared in assigning judgements	▪ providing a correct network structure amongst criteria is difficult even for experts, and different structures lead to different results ▪ the formation of a super matrix requires pairwise comparison of all criteria with all other criteria, which is a step that is difficult and unnatural
AHP	▪ allows decision makers to structure the decision-making problem into hierarchy trees ▪ makes the problem easy to understand	▪ time-consuming due to the number of pairwise comparisons ▪ requires mathematical calculations, which increase as the number of criteria and alternatives increase or change in terms of scoring used as the scoring in AHP, which relies on the alternatives considered for evaluation ▪ addition and deletion of alternatives may affect the final ranking
TOPSIS	▪ significant approach to solve real-world problems ▪ application in discretising alternative challenges ▪ can immediately recognise proper alternatives ▪ decreases in the number of required pairwise comparisons, with a capacity limitation that does not necessarily control the process ▪ useful when alternatives and attributes are numerous and when quantitative or objective data are available ▪ basis in aggregating function representing 'closeness to the ideal,' which originates from compromise programming method	▪ TOPSIS includes the lack of provision to weigh elicitation and check the consistency of judgements

forms. Lastly, three of the experts help us obtain the weights for evaluation criteria by filling up the BWM standard.

b: CHECKLIST FORM FOR EVALUATION

Similar to many studies (e.g. [54], [55]), a checklist form is employed in this study. This easy-to-use form is prepared in English language. The checklist is used to evaluate the alternatives (apps) with respect to the criteria (LSRW) (see Appendix B) and comprises 21 questions, which are divided into three sections that measure the LSRW skills.

Before the respondents evaluate the actual apps, the checklist is developed and reviewed by the experts to: (1) check whether the criteria are suitable for evaluating the English learning apps in early childhood environment, (2) identify the question problems, (3) breakdown the question-answering process (4) and determine other potential measurement errors in our checklist form [55], [56].

c: EVALUATION PROCEDURES OF MOBILE APPS FOR ENGLISH LEARNING BASED ON LSRW SKILL CRITERIA

This study is conducted using a checklist form, which is administered to evaluate the alternatives (apps). Each alternative is evaluated with respect to all identified criteria. Initially, all the apps are installed on a tablet device (with an Android operating system), and all the installed apps are subsequently presented to the experts for evaluation. The form is distributed for the respondents to check the criterion suitability of early childhood English learning apps and then answer the evaluation questions. To eliminate bias in this study, the forms

are distributed and collected without the intervention of any other person, and the respondents should answer the form questions. The forms are manually distributed and collected, not electronically, to maintain confidentiality. The collected evaluation data in the checklist form are analysed by applying MCDM TOPSIS method, which is required for a comprehensive ranking of English learning mobile apps.

B. BENCHMARKING PHASE

This phase aims to benchmark young learners' English learning mobile apps on the basis of the identified DM through MCDM. Various MCDM theories are discovered. The most common and famous MCDM methods that use different concepts comprise multiplicative exponential weighting (MEW), weighted product method (WPM), weighted sum model (WSM), analytic network process (ANP), analytic hierarchy – process (AHP), BWM and TOPSIS. Our related literature analysis implies that these methods have not been applied yet to evaluate and benchmark mobile apps in an English learning environment, which is considered a theoretical gap.

Studies of [57]–[60] explain the weaknesses and strengths of common MCDM methods, including recommendations. Table 2 summaries the explanation.

The latest trend in the field entails the integration of two or more decision making methods [61]–[63] to prevent shortcomings in a single method. Therefore, TOPSIS needs an effective approach to assign the relative importance of various criteria with respect to the goal. AHP and
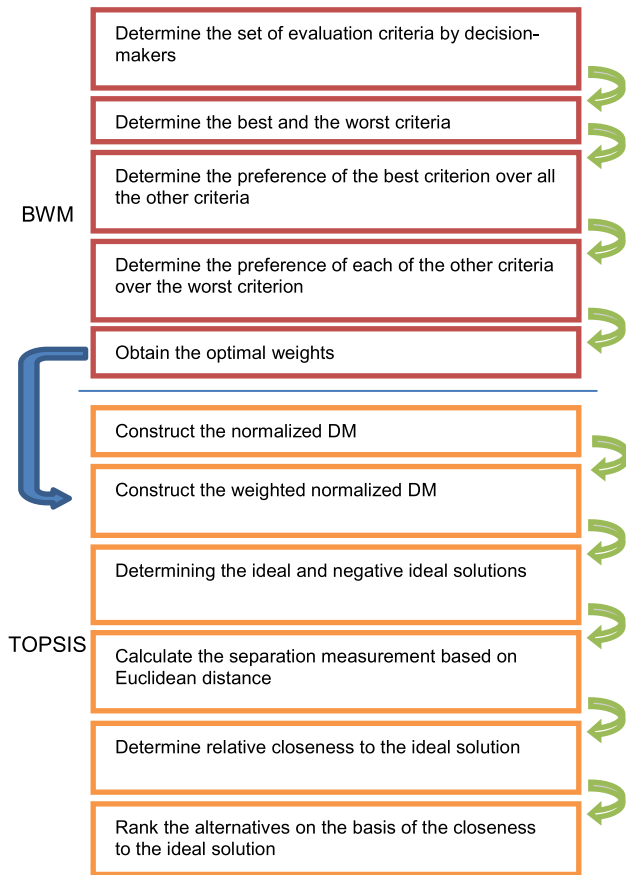
IEEE *Access*

N. K. Ibrahim *et al.*: Multi-Criteria Evaluation and Benchmarking for Young Learners' English Language Mobile Applications

BWM
Determine the set of evaluation criteria by decision-makers
Determine the best and the worst criteria
Determine the preference of the best criterion over all the other criteria
Determine the preference of each of the other criteria over the worst criterion
Obtain the optimal weights

TOPSIS
Construct the normalized DM
Construct the weighted normalized DM
Determining the ideal and negative ideal solutions
Calculate the separation measurement based on Euclidean distance
Determine relative closeness to the ideal solution
Rank the alternatives on the basis of the closeness to the ideal solution

FIGURE 3. Integrated BWM–TOPSIS model for app selection.

BWM offer a procedure to obtain the relative importance of different criteria. These methods are used to assign weights for criteria depending on experts' opinions [64]–[68]. Statistical results reveal that BWM is more accurate and consistent than AHP. The remarkable features of BWM, compared with AHP method, include requiring only a few pairwise comparisons. This requirement leads to consistent comparisons, thereby providing reliable results [65], [69].

Recently, the BWM–TOPSIS methods have become a widely accepted integrated multi-criteria decision analysis method (e.g. [69]–[72]). These proposed integrated BWM–TOPSIS methods have several merits. Firstly, the newest MCDM method, namely, BWM is applied to determine criterion weight, which is easier to operate than other methods. Secondly, the most common MCDM method [73], namely, TOPSIS is employed for benchmarking and ranking English learning apps; this ranking procedure is easier and clearer to implement than other MCDM methods [71]. Thus, in this study, BWM was used to assign and determine the criteria weights, whereas TOPSIS was employed to benchmark English learning apps. Table 5 presents the weighting and ranking processes of these two methods in the constructed DM. BWM and TOPSIS steps are explained in Fig. 3 and are further discussed afterward.

1) UTILISING BWM METHOD FOR CRITERIA WEIGHT DETERMINATION

BWM assigns weights for a set of evaluation criteria through pairwise comparison of the best (most desirable or most important) and the worst (least desirable or least important) criteria with the other criteria in the evaluation criteria set [74]. BWM form, which is designed according to the BWM steps, is used to obtain weights of the aspects (LSRW) and their criteria (see Appendix A). All aspects and their criteria are weighted by the same selected panel of experts as mentioned in the 'Identification' phase. In this stage, several steps are involved to assign proper weights to the criteria by using BWM. The BWM procedure includes the following steps [68], [71], [75].

a: DETERMINE THE SET OF EVALUATION CRITERIA BY DECISION MAKERS

The first step of the BWM is determining the criteria set; {c1, c2, …, cn} should be used by the decision maker when deciding for the best alternative. In this study, the criteria set is obtained from the 2016 KSPK standard as mentioned in the 'Identification' phase.

b: DETERMINE THE BEST AND WORST CRITERIA

In this step, decision makers choose the best and worst criteria amongst the set of criteria identified in the previous step from their perspective. The best criteria represent the most important criteria, whereas the worst criteria are the least important criteria considered for the decision.

c: DETERMINE THE PREFERENCE OF THE BEST CRITERION OVER ALL THE OTHER CRITERIA

The process of pairwise comparison is conducted between the identified best criterion and the other criteria. This step aims to determine the preference of the best criterion over all the other criteria. The experts choose a value from 1 to 9 (1 = equally important, 9 = extremely important) to represent the importance of the best criterion over the other criteria. This procedure results in a vector, namely, 'Best-to-Others,' which can be: $AB = (aB1, aB2, …, aBn)$, where aBi indicates the importance of the best criterion B over criterion j, and $aBB = 1$.

d: DETERMINE THE PREFERENCE OF EACH OF THE OTHER CRITERIA OVER THE WORST CRITERION

This pairwise comparison aims to identify the preference of all criteria over the least important criterion. The evaluator/expert determines the importance of all the criteria over the worst criterion; numbers from 1 to 9 indicate the level of importance. The result of this step is a vector, namely, 'Others-to-Worst.' Its result is represented as $Aw = (a1w, a2w, …, anw)$, where ajw represents the preference of criterion j over the worst criterion w. Evidently, $aww = 1$. Fig. 4 clarifies the two types of reference comparisons, known as Best-to-Others and Others-to-Worst criteria.
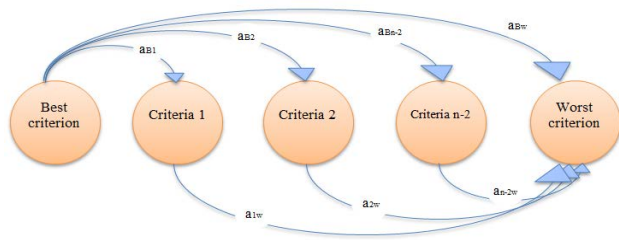
N. K. Ibrahim *et al.*: Multi-Criteria Evaluation and Benchmarking for Young Learners' English Language Mobile Applications

IEEE *Access*

FIGURE 4. Reference comparisons in the BWM method.

e: OBTAIN THE OPTIMAL WEIGHTS ($W*1, W*2, \ldots, W*n$)

The optimal weight for the criteria is the one wherein for each pair of *WB/Wj* and *Wj/Ww*, we have $WB/Wj = aBj$ and $Wj/Ww = ajw$. To meet these conditions for all *j*, the maximum absolute differences *{|WB - aBjWj|* and *|Wj - ajwWw|}* for all *j* are minimised. Considering the non-negativity and sum condition for the weights, the following problem is created.

$$min\ max\ j\ \left\{\left|\frac{W_B}{W_j} - a_{Bj}\right|, \left|\frac{W_j}{W_w} - a_{jw}\right|\right\}$$
$$\sum_{j}^{s.t.} wj = 1, \quad wj \geq 0\ for\ all\ j \quad (1)$$

This model can be solved by transforming it to the linear programming formulation.

$$min\ \xi$$
$$s.t.\ \left|\frac{W_B}{W_j} - a_{Bj}\right| \leq \xi, \quad for\ all\ j \quad (2)$$
$$\left|\frac{W_j}{W_w} - a_{jw}\right| \leq \xi, \quad for\ all\ j \quad (3)$$
$$\sum_{j} W_j = 1$$
$$W_j \geq 0, \quad for\ all\ j$$

By solving this problem, the optimal weights *(W*1, W*2, \ldots, W*n)* and the optimal value of ξ^* are acquired. ξ^* refers to the outcomes' reliability, which depends on the extent of consistency in the comparisons. Thus, the closer ξ^* is to a zero value, the more consistent the comparison system provided by the decision makers. Subsequently, the consistency ratio (CR) is calculated using ξ^*, and the corresponding consistency index is as follows:

$$CR = \xi^*/consistency\ index. \quad (4)$$

Table 3 depicts the consistency index. It shows that the lower the CR, the higher the reliability of the comparisons. That is, a CR value of less than 0.1 is reflective of consistent judgements [76].

2) UTILISING TOPSIS METHOD TO BENCHMARK AND RANK ENGLISH LEARNING APPS

TOPSIS is based on the idea of an alternative with respect to the closest to the ideal solution and the farthest from the

TABLE 3. Index of consistency.

a_{BW}	1	2	3	4	5	6	7	8	9
Consistency Index	0.0	0.44	1.0	1.63	2.30	3.00	3.73	4.47	5.23

anti-ideal solution as the best option [77]. In this section, TOPSIS, which is considered a proper method amongst MCDM techniques, is utilised. This method involves several steps to implement. The procedures of the TOPSIS method are demonstrated in the subsequent steps [76], [78].

a: CONSTRUCT THE NORMALISED DM

This process attempts to transform the various attribute dimensions into non-dimensional attributes and allows a comparison amongst the attributes. The matrix is $(Xij)m*n$, then the form $(Xij)m*n$ is normalised to the matrix $R = (rij)m*n$ by using the normalisation method:

$$r_{ij} = x_{ij}/\sqrt{\sum_{i=1}^{m} x_{ij}^2} \quad (5)$$

This process results in a new matrix R, which is expressed as follows:

$$R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ r_{m1} & r_{m2} & \cdots & r_{mn} \end{bmatrix}$$

b: CONSTRUCT THE WEIGHTED NORMALISED DM

In this stage, a set of weights, $W = W*1, W*2, W*3, \ldots, W*n$, from the decision maker is accommodated to the normalised DM. The resulting matrix can be calculated by multiplying each column from normalised DM (*R*) with its associated weight *Wj*. Notably, the set of the weights should be equal to 1.

$$\sum_{j=1}^{m} w_j = 1 \quad (6)$$

This process can result in a new matrix *V*, which is expressed as follows:

$$v = \begin{bmatrix} v_{11} & v_{12} & \cdots & v_{1n} \\ v_{21} & v_{22} & \cdots & v_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ v_{m1} & v_{m2} & \cdots & v_{mn} \end{bmatrix}$$
$$= \begin{bmatrix} w_1 r_{11} & w_2 r_{12} & \cdots & w_n r_{1n} \\ w_1 r_{21} & w_2 r_{22} & \cdots & w_n r_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ w_1 r_{m1} & w_2 r_{m2} & \cdots & w_n r_{mn} \end{bmatrix}$$

IEEE *Access*

N. K. Ibrahim *et al.*: Multi-Criteria Evaluation and Benchmarking for Young Learners' English Language Mobile Applications

c: DETERMINE THE POSITIVE AND NEGATIVE IDEAL SOLUTIONS

In this process, two alternatives, A^* (positive ideal alternative) and $A-$ (negative ideal alternative), are defined as follows:

$$A^* = \left\{ \left(\left(\max_i v_{ij} | j \in J \right), \left(\min_i v_{ij} | j \in J^- \right) \right| \right.$$
$$\left. i = 1, 2, \ldots, m \right) \right\}$$
$$= \left\{ v_1^*, v_2^*, \ldots, v_j^*, \cdots v_n^* \right\} \qquad (7)$$
$$A^- = \left\{ \left(\left(\min_i v_{ij} | j \in J \right), \left(\max_i v_{ij} | j \in J^- \right) | i = 1, 2, \ldots, m \right) \right\}$$
$$= \left\{ v_1^-, v_2^-, \ldots, v_j^-, \cdots v_n^- \right\} \qquad (8)$$

Note that J is a subset of $\{i = 1, 2, \ldots, m\}$, which presents the benefit attribute, whereas $J-$ is the complement set of J or (Jc), which is the set of cost attribute.

d: CALCULATE THE SEPARATION MEASUREMENT ON THE BASIS OF THE EUCLIDEAN DISTANCE

In this stage, the separation measurement is performed by calculating the distance between each alternative in V and the ideal vector A^* by applying the Euclidean distance, which is expressed as:

$$S_{i^*} = \sqrt{\sum_{j=1}^n \left(v_{ij} - v_j^* \right)^2}, \quad i = (1, 2, \cdots m) \qquad (9)$$

Similarly, the separation measurement for each alternative in V from the negative ideal $A-$ is expressed as follows:

$$S_{i^-} = \sqrt{\sum_{j=1}^n \left(v_{ij} - v_j^- \right)^2}, \quad i = (1, 2, \cdots m) \qquad (10)$$

As a result of this step, two values, namely, Si^* and Si- for each alternative are counted, and both represent the distance between each alternative and the positive and negative ideal solutions.

e: DETERMINE RELATIVE CLOSENESS TO THE IDEAL SOLUTION

The closeness of Ai to the ideal solution A^* is defined as:

$$C_{i^*} = S_{i^-} / \left(S_{i^-} + S_{i^*} \right), \quad 0 < C_{i^*} < 1,$$
$$i = (1, 2, \cdots m). \qquad (11)$$

$Ci^* = 1$ if and only if $(Ai = A^*)$. Similarly, $Ci^* = 0$ if and only if $(Ai = A)$.

f: RANK THE ALTERNATIVES ON THE BASIS OF THE CLOSENESS TO THE IDEAL SOLUTION

The set of alternative Ai can be ranked in the descending order of Ci^*, where a high value is preferred. Two contexts can be adopted in MCDM to rank the alternatives, namely, individual and group decision making. The individual context refers to ranking the alternatives with respect to each expert's perspective, which means the three rank results of Ci^* for the three experts are present. The group decision-making context

has two approaches. The first one is internal aggregation that is calculated through dividing the summation values of the negative separation by the negative separation values plus the positive separation values for each expert.

$$Internal\ aggregation = sum(S_i -)/(sum(S_i -) + sum(S_i^*)). \qquad (12)$$

The second approach is external aggregation, which is calculated by finding the average ranking values for each expert.

IV. RESULTS AND DISCUSSION

This section presents the results of the proposed DM for evaluating and benchmarking the English learning apps in terms of LSRW skills based on the 2016 KSPK standard.

A. DATA PRESENTATIONIN THE DM

This section presents the results obtained from the evaluation process for the six English learning apps with respect to LSRW criteria (see 'Identification' phase). Table 4 presents the results obtained from the first expert, and Appendix C (Tables 26–30) provide detailed results of the five other experts. Given the differences in the expert evaluation of English learning apps, the evaluation average of the six experts is calculated. Table 5 presents the completed DM. The table shows that the evaluation of each English learning app is based on 17 criteria. The following section discusses the result of the integration between BWM and TOPSIS methods.

B. RESULTS OF BENCHMARKING ENGLISH LEARNING MOBILE APPS FOR YOUNG LEARNERS BASED ON THE IDENTIFIED DM USING BWM AND TOPSIS METHODS

This section presents the results in two subsections. The first subsection is the criteria weighting results by using the BWM method, followed by the second subsection, which is related to the results of the TOPSIS method for ranking.

1) RESULTS FOR WEIGHT USING BWM METHOD

This section presents and explains the BWM results. Only three experts are asked for their preferences on the evaluation criteria of English learning apps via a BWM comparison question form. Table 6 presents the first expert results of the main criteria and their sub-criteria, and the detailed results of the two other experts are shown in Appendix C (Table 31).

To calculate the criteria weights according to the BWM method, the best and worst criteria are identified, the comparison between the best criterion and the others is performed, and the comparison between all criteria with worst criterion is achieved. Finally, the linear model of BWM is solved according to Eqs. (1), (2) and (3) to obtain the weights, and Eq (4) is used to calculate the CR of each expert's preference, as cited in the 'Benchmarking' phase.

To calculate the global weights of each criterion for the three experts, the BWM method derives the local weights for each criterion at each level, as shown in Table 6 and

N. K. Ibrahim *et al.*: Multi-Criteria Evaluation and Benchmarking for Young Learners' English Language Mobile Applications

IEEE *Access*

TABLE 4. Results obtained from the first expert.

Main Criteria Sub-criteria (apps)	Listening and Speaking									Reading				Writing			
	Stimulus given (sg)	Rimes (ri)	Poems and rhymes (p&r)	Stories (st)	Favourite things and activities (ft&a)	Oral texts (ot)	Familiar activities and experiences (fa&e)	Stories heard (sh)	Daily situations (ds)	Alphabet letters (al)	Simple phrases (sp)	Simple sentences (ss)	Text reading (tr)	Copy legible phrases (clp)	Copy legible sentences (cls)	Idea and information communication (i&ic)	Legible writing (lw)
Lingokids	1	1	1	1	1	1	1	0	1	1	0	0	1	1	1	1	1
Fun English	1	0	0	0	1	1	1	0	1	1	0	1	1	1	0	1	1
FunWithFlupe	1	0	0	0	1	0	1	0	0	1	1	1	1	1	1	1	1
First Words	1	0	0	0	1	1	1	0	1	1	0	0	1	1	0	1	1
Montessori	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Spelling Bee	1	0	0	0	1	1	1	0	1	1	1	0	1	1	1	1	1

TABLE 5. Completed DM.

Main Criteria Sub-criteria (apps)	Listening and Speaking									Reading				Writing			
	Stimulus given (sg)	Rimes (ri)	Poems and rhymes (p&r)	Stories (st)	Favourite things and activities (ft&a)	Oral texts (ot)	Familiar activities and experiences (fa&e)	Stories heard (sh)	Daily situations (ds)	Alphabet letters (al)	Simple phrases (sp)	Simple sentences (ss)	Text reading (tr)	Copy legible phrases (clp)	Copy legible sentences (cls)	Idea and information communication (i&ic)	Legible writing (lw)
Lingokids	0.6667	1	0.8333	0.8333	0.6667	0.5	0.8333	0.3333	0.5	0.5	0.3333	0.3333	0.5	0.6667	1	1	0.6667
Fun English	1	0.6667	0.6667	0.1667	1	0.6667	0.5	0.3333	0.8333	0.6667	0.3333	0.8333	0.6667	0.6667	0.1667	0.6667	0.6667
FunWith-Flupe	0.8333	0	0.5	0.5	0.5	0.3333	0.8333	0	0	0.8333	1	0.8333	0.8333	0.8333	1	1	1
First Words	0.8333	0.3333	0	0.3333	0.8333	0.8333	1	0.1667	0.6667	0.8333	0.1667	0.1667	0.6667	0.8333	0	0.8333	0.8333
Montessori	1	0.5	1	1	1	0.8333	0.8333	1	0.8333	0.5	1	0.8333	1	0.8333	0.8333	0.8333	1
Spelling Bee	1	0.1667	0.3333	0.5	0.5	0.8333	1	0	0.8333	0.8333	1	0	1	0.6667	0.8333	0.6667	1

TABLE 6. Results of BWM method for the weight preferences of the criteria of english learning app evaluation (first expert).

	Expert 1						
	Main Criteria						
List of criteria	Best criterion	Other Criteria	Scores	Other criteria	Worst criterion	Scores	Weight
Listening and speaking	Listening and speaking	Reading	9	Listening and speaking	reading	9	0.750
Reading		Writing	5	writing		6	0.062
Writing		....	.....	....		....	0.188
Consistency: 0.036							
	Subcriteria of listening and speaking						
List of criteria	Best criterion	Other Criteria	Scores	Other criteria	Worst criterion	Scores	Weight
Stimulus given	Daily situations	Stimulus given	2	Stimulus given	Poems and rhymes	3	0.152
Rimes		Rimes	6	Rimes		5	0.065
Poems and rhymes		Poems and rhymes	9	Stories		5	0.027
Stories		Stories	3	Favourite things and activities		5	0.129
Favourite things and activities		Favourite things and activities	5	Oral texts		4	0.077
Oral texts		Oral texts	4	Familiar activities and experiences		3	0.097
Familiar activities and experiences		Familiar activities and experiences	5	Stories heard		2	0.077
Stories heard		Stories heard	4	Daily situations		9	0.061
Daily situations		....	.....	....		.....	0.315
Consistency: 0.014							
	Subcriteria of reading						
List of criteria	Best criterion	Other Criteria	Scores	Other criteria	Worst criterion	Scores	Weight
Alphabet letters	Alphabet letters	Simple phrases	2	Alphabet letters	Text reading	7	0.496
Simple phrases		Simple sentences	4	Simple phrases		3	0.294
Simple sentences				Simple sentences		6	0.160
Text reading		Text reading	7				0.050
Consistency: 0.038							
	Subcriteria of writing						
List of criteria	Best criterion	Other Criteria	Scores	Other criteria	Worst criterion	Scores	Weight
Copy legible phrases	Idea and information communication	Copy legible phrases	3	Copy legible phrases	Legible writing	4	0.195
Copy legible sentences		Copy legible sentences	3	Copy legible sentences		4	0.195
Idea and information communication		Legible writing	9	Idea and information communication		9	0.553
Legible writing		....	.....	....		......	0.057
Consistency: 0.007							

Appendix C (Table 31), representing the importance of each criterion regarding the parent. The global weight for each criterion is obtained, representing the importance of that criterion with respect to the goal for each expert.

The weight of each criterion is determined by comparing the criteria on the basis of the BWM. These weights are called local weights. The weights of the original criteria and their associated local weights are multiplied to determine the

N. K. Ibrahim et al.: Multi-Criteria Evaluation and Benchmarking for Young Learners' English Language Mobile Applications

IEEE Access

global weights with respect to experts' goals, as presented in Table 7. In addition, Table 7 presents the overall local and global weights obtained from the three experts for the 17 evaluation criteria. In addition, the table shows that the overall CR for the comparison according to each expert's scores is an acceptable ratio of less than 0.1, which reflects the high consistency of the comparison outcomes, as mentioned in the 'Benchmarking' phase. The global weights are used in the proposed DM because they represent the importance of the criteria with respect to the goal.

Table 6 indicates that the first expert assigns the maximum weight for 'daily situations' as 0.236, and the minimum weight obtained by 'text reading' is 0.003. The second expert assigns the maximum weight for 'idea and information communication' criterion as 0.275, and the minimum weight obtained by 'text reading' is 0.005. The third expert assigns the maximum weight for 'idea and information communication' as 0.204, and the minimum weight obtained by 'simple phrases' is 0.011. The final weight results are used in applying the TOPSIS method in the subsequent section, as mentioned in the 'Benchmarking' phase.

2) RANKING RESULTS OF TOPSIS METHOD

This section discusses the ranking results of the English learning apps according to the weighted evaluation criteria. Two main decision-making contexts are included in this study, namely, individual and group. Furthermore, two approaches are used in the group context, namely, internal and external aggregations.

a: TOPSIS RESULTS OF INDIVIDUAL CONTEXT FOR DIFFERENT WEIGHTS ACCORDING TO EXPERTS

TOPSIS is used to rank the alternatives on the basis of the DM results, as presented in Table 5. The results presented in Table 7 indicate the global importance of the evaluation criteria from the viewpoint of each expert. As described in the 'Benchmarking' phase, the TOPSIS technique depends on comparing each alternative with ideal solutions. S− and S* represent the closeness of an alternative to the negative and positive ideal solutions, respectively. Table 8 shows the TOPSIS ranking results on the basis of the weights that reflect the viewpoint of the first expert. The two remaining experts' TOPSIS results are shown in Appendix C (Table 32). Figure 5 illustrates the virtualised overall TOPSIS final ranking results on the basis of the three experts' preferences. The three rank results indicate the highest rank values of 0.8689, 0.7464 and 0.7845 for the Montessori app. In addition, the three rank results reveal that the lowest values are 0.2193, 0.3355 and 0.3696 for the FunWithFlupe, Fun English and First Words apps, respectively.

Considering the previous discussion, the results of the individual context clearly show variation amongst the rankings of the three experts. Thus, a group TOPSIS decision-making context must be applied to provide an alternative ranking that

TABLE 7. Local and global weights for the three experts.

First expert				
Main criteria	Weights	Subcriteria	Local weights	Global weights
Listening and speaking	0.750	Stimulus given	0.152	0.114
		Rimes	0.065	0.049
		Poems and rhymes	0.027	0.020
		Stories	0.129	0.097
		Favourite things and activities	0.077	0.058
		Oral texts	0.097	0.073
		Familiar activities and experiences	0.077	0.058
		Stories heard	0.061	0.046
		Daily situations	0.315	0.236
Reading	0.062	Alphabet letters	0.496	0.031
		Simple phrases	0.294	0.018
		Simple sentences	0.160	0.010
		Text reading	0.050	0.003
Writing	0.188	Copy legible phrases	0.195	0.037
		Copy legible sentences	0.195	0.037
		Idea and information communication	0.553	0.104
		Legible writing	0.057	0.011
Overall CR < 0.1				
Second expert				
Main criteria	Weights	Subcriteria	Local weights	Global weights
Listening and speaking	0.325	Stimulus given	0.062	0.020
		Rimes	0.024	0.008
		Poems and rhymes	0.051	0.017
		Stories	0.154	0.050
		Favourite things and activities	0.250	0.081
		Oral texts	0.048	0.016
		Familiar activities and experiences	0.154	0.050
		Stories heard	0.103	0.033
		Daily situations	0.154	0.050
Reading	0.100	Alphabet letters	0.564	0.056
		Simple phrases	0.164	0.016
		Simple sentences	0.220	0.022
		Text reading	0.052	0.005
Writing	0.575	Copy legible phrases	0.254	0.146
		Copy legible sentences	0.193	0.111
		Idea and information communication	0.478	0.275
		Legible writing	0.075	0.043
Overall CR < 0.1				
Third expert				
Main criteria	Weights	Subcriteria	Local weights	Global weights
Listening and speaking	0.542	Stimulus given	0.070	0.038
		Rimes	0.095	0.051
		Poems and rhymes	0.036	0.020
		Stories	0.230	0.125
		Favourite things and activities	0.140	0.076
		Oral texts	0.122	0.066
		Familiar activities and experiences	0.045	0.024
		Stories heard	0.122	0.066
		Daily situations	0.140	0.076
Reading	0.166	Alphabet letters	0.518	0.086
		Simple phrases	0.069	0.011
		Simple sentences	0.241	0.040
		Text reading	0.172	0.029
Writing	0.292	Copy legible phrases	0.117	0.034
		Copy legible sentences	0.134	0.039
		Idea and information communication	0.698	0.204
		Legible writing	0.051	0.015
Overall CR < 0.1				

IEEE *Access*

N. K. Ibrahim *et al.*: Multi-Criteria Evaluation and Benchmarking for Young Learners' English Language Mobile Applications

TABLE 8. Ranking results based on the first expert's weights.

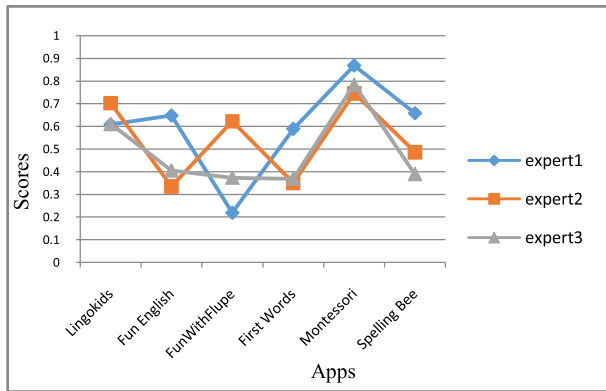| NO.	App	S-	S*	Rank value	Final rank
1)	Lingokids	0.0963	0.0619	0.6087	4
2)	Fun English	0.1247	0.0677	0.6480	3
3)	FunWithFlupe	0.0385	0.1370	0.2193	6
4)	First Words	0.1008	0.0702	0.5895	5
5)	Montessori	0.1430	0.0216	0.8689	1
6)	Spelling Bee	0.1256	0.0652	0.6583	2

FIGURE 5. Ranking results based on the three experts' weights.

TABLE 9. Group decision making of TOPSIS with internal and external aggregations.

App	S-	S*	Internal aggreg- ation	Intern- al rank	External aggregate- on	Extern- al rank
Lingokids	0.2621	0.1500	0.6360	2	0.6409	2
Fun English	0.2249	0.2347	0.4894	4	0.4632	4
FunWith- Flupe	0.1743	0.2792	0.3843	6	0.4055	6
First Words	0.1915	0.2306	0.4537	5	0.4367	5
Montesso ri	0.3325	0.0788	0.8084	1	0.7999	1
Spelling Bee	0.2439	0.2195	0.5264	3	0.5117	3

considers all decision makers. The following section presents the results of the group TOPSIS decision-making context.

b: GROUP TOPSIS WITH INTERNAL AND EXTERNAL AGGREGATIONS

To extend TOPSIS to a group decision environment, two approaches—internal and external aggregations—are reported in the literature, as mentioned in the 'Benchmarking' phase. Table 9 presents the results of the alternatives of group TOPSIS with internal and external aggregations.

The virtualised results in Figures 6 and 7 for the internal and external aggregation rankings, respectively, indicate similar ranks using the aforementioned methods. Henceforth, the findings of the external aggregation method is considered
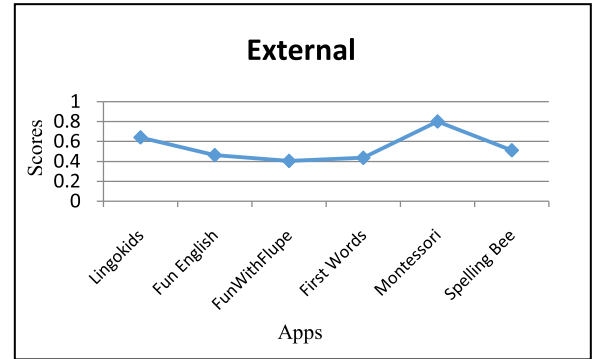

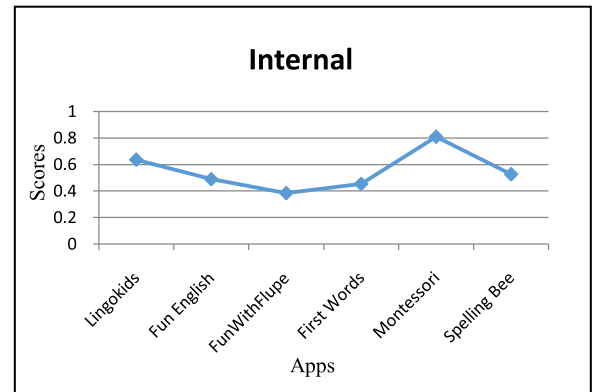**FIGURE 6.** Results of external aggregation.

FIGURE 7. Results of internal aggregation.

the final ranking results and are used in the validation processes, similar to the study of [79]. The subsequent section describes the external aggregation validation results in detail.

V. VALIDATION AND EVALUATION
A. VALIDATION
1) OBJECTIVE VALIDATION

The statistical method (mean) is utilised to ensure the systematic ranking of the English learning apps. The mean refers to the average, which is calculated by dividing the sum of the observed results by the number of results, as shown as follows:

$$\bar{x} = \sum_{i=1}^{n} x_i / n \qquad (13)$$

The mean is used to ensure the validity and systematic ranking of the proposed DM results. The scoring of the six apps is divided into three groups on the basis of the ranked result based on the TOPSIS method, which is similar to the study [72]. The results are expressed as the mean for each group.

The validation process must prove that the first group has the highest scoring value by calculating the mean and comparing it with that of the other groups. The mean of the second group must be lower than that of the first group. Meanwhile, the mean of the third group must be lower than that of the first and second groups. The systematic ranking results show that

N. K. Ibrahim et al.: Multi-Criteria Evaluation and Benchmarking for Young Learners' English Language Mobile Applications

IEEE Access

TABLE 10. Validation results of external group decision - making ranking.

	Stimulus given	Rimes	Poems and rhymes	Stories	Favourite things and activities	Oral texts	Familiar activities and experiences	Stories heard	Daily situations	Alphabet letters	Simple phrases	Simple sentences	Text reading	Copy legible phrases	Copy legible sentences	Idea and information communication	Legible writing
	Validation results for the external aggregation group decision making																
App5	0.0261	0.0133	0.0120	0.0594	0.0376	0.0253	0.0176	0.0432	0.0603	0.0166	0.0083	0.0134	0.0063	0.0326	0.0281	0.0783	0.0107
App1	0.0174	0.0266	0.0100	0.0495	0.0250	0.0152	0.0176	0.0144	0.0362	0.0166	0.0028	0.0054	0.0031	0.0261	0.0337	0.0940	0.0072
Overall mean	0.0262																
App6	0.0261	0.0044	0.0040	0.0297	0.0188	0.0253	0.0211	0.0000	0.0603	0.0276	0.0083	0.0000	0.0063	0.0261	0.0281	0.0626	0.0107
App2	0.0261	0.0177	0.0080	0.0099	0.0376	0.0203	0.0106	0.0144	0.0603	0.0221	0.0028	0.0134	0.0042	0.0261	0.0056	0.0626	0.0072
Overall mean	0.0208																
App4	0.0217	0.0089	0.0000	0.0198	0.0313	0.0253	0.0211	0.0072	0.0483	0.0276	0.0014	0.0027	0.0042	0.0326	0.0000	0.0783	0.0090
App3	0.0217	0.0000	0.0060	0.0297	0.0188	0.0101	0.0176	0.0000	0.0000	0.0276	0.0083	0.0134	0.0052	0.0326	0.0337	0.0940	0.0107
Overall mean	0.0197																

the first group should be statistically proven as the highest amongst all groups [68], [80].

2) VALIDATION RESULT

This section presents the validation processes of external group decision-making rankings. This study employs objective validation processes. The validation process for the ranking results of English learning apps is performed by dividing the ranking result into three equal groups with two apps each. Mean is calculated for each group to ensure the systematic ranking of English learning apps. After the normalisation and weighting processes for the data of the first, second and third groups of English learning apps, Table 10 presents the validation results for the external aggregation group decision making. The mean value in the first group (0.0262) is higher than the mean values in the second and third groups (0.0208 and 0.0197, respectively). The mean value in the second group (0.0208) is higher than in the third group (0.0197). Thus, the internal and external group decision-making rankings are objectively valid and systematic.

B. EVALUATION

The most relevant existing studies related to English learning app evaluation and benchmarking are found in [9], [14], [36]–[38], [40], as shown in the Literature review section. In the current section, the proposed DM is evaluated and compared with these six relevant studies (benchmark studies). Comparison requires the provision of scenarios and a benchmarking checklist. Each scenario reflects issues that must be defined and addressed in the studies on evaluating and benchmarking English learning apps. These issues represent the points of comparison for the proposed DM with the benchmark studies in the checklist. A benchmarking checklist provides a useful way to measure how effective the proposed work is compared with other works. The comparisons are performed on the basis of whether the compared works cover the issues addressed in the comparison scenario, as in studies [79], [81]. Three scenarios are clarified as follows to show the comparison points in the benchmarking checklist.

In the first scenario, the comparison between the proposed and benchmark studies is based on app evaluation and related comparison points. In several studies, the evaluation process of apps should be conducted on the basis of content evaluation and usability evaluation.

During content evaluation, the learning of any language involves and focuses on the mastery of LSRW skills [82], [83]. Thus, app content should be evaluated in terms of the four skills. The first skill is listening, which is one of the most crucial language skills [84], [85]. Effective teaching methods for English learning must begin with this skill [10]. However, students believe that speaking is the most important language skill that should be mastered. They argue that learning achievement must be assessed on the basis of speaking skill [86].

Moreover, amongst the LSRW skills, reading plays an essential role in understanding and learning authentic materials. In the context of English as a foreign language, people do not often have many chances to interact and communicate orally with native speakers. Thus, reading can play an important role in learning improvement [87]. In general, writing is the last language skill to be gained for language learners.

IEEE Access

N. K. Ibrahim *et al.*: Multi-Criteria Evaluation and Benchmarking for Young Learners' English Language Mobile Applications

Success in English writing brings learners benefits not only in their English learning but also in their entire life [10], [88].

Along with LSRW skills, vocabulary mastery is a basic factor for English learning. Mastery of vocabulary is important for anyone learning the language. Foreign language learners can speak fluently and accurately, write easily or understand what they read or hear when they have enough vocabulary and can use it accurately [89].

In relation to usability evaluation, layout design is important to encourage learners to use apps, as such a design represents the first element seen by users when opening apps [90]. Raising motivation is the key in any kind of learning. Accordingly, students with high motivation can achieve further English learning [91]. Moreover, if an app is easy to use and useful, users can have a positive attitude towards it, which in turn increases their intention to use the app. Ease of use is related to the degree to which a person believes that using the app is free of effort, whereas usefulness refers to the degree to which the learner believes that using the app can enhance his or her learning [92].

In the second scenario, compression is done on the basis of the benchmarking procedure of English learning apps and related comparison points. Benchmarking is performed to compare English learning apps under the same conditions [93]. Several criteria influence the benchmarking process. All criteria should be considered for assessment [19]. Consequently, different weights are generally given for the criteria. Thus, criteria weighting is significant [94] and represents a key objective by benchmarking [17].

Moreover, data variation amongst different criteria during benchmarking is becoming a major challenge because measuring the alternatives in terms of criteria can be represented as a set of values [28], [29]. Data variation amongst these values causes a problem in which decision makers cannot compare an alternative with others [31], [32].

In the third scenario, compression is performed on the basis of the validation and evaluation of the proposed and benchmark works. Validation is the process of checking whether a proposed work is valid and appropriate for its purpose [79]. Evaluation is related to the process of comparing the performance and accuracy of the proposed work [81].

After detailing the comparison scenarios, several comparison points are recognised and highlighted for each scenario that must be considered in English learning app evaluation and benchmarking. Comparison points are extracted, and Figure 8 describes the connection between scenarios and their related issues, which are defined as points of comparison in the benchmarking checklist. The descriptions of the checklist comparison points are presented as follows:

- **Listening skill**: Listening is the process of understanding speech. Out the four skills LSRW, listening is the most important [85], [95]. Thus, this point of comparison is included in the benchmarking checklist to demonstrate whether the listening skill evaluation has been provided in the study.
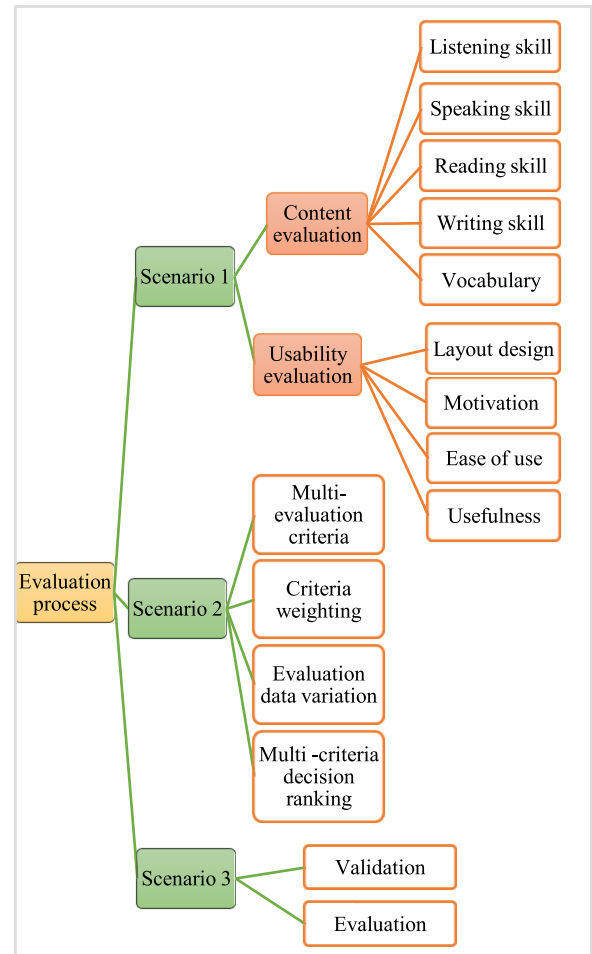

FIGURE 8. Relationships amongst comparison points and scenarios.

- **Speaking skill**: The speaking skill is an important part of language learning and teaching [96]. This issue is included because it plays an important role in giving learners the ability to communicate in English [97].
- **Reading skill**: The importance of reading skill is prominent [98]. Thus, this point of comparison is included in the benchmarking checklist to demonstrate whether the reading skill evaluation has been provided in the study.
- **Writing skill**: Writing represents one method of expressing thoughts. Therefore, this skill is very important [99] and is included as a point in the benchmarking checklist.
- **Vocabulary**: This point indicates whether the study has provided an evaluation of vocabulary issues. Vocabulary is considered the foundation of English language learning. Having ample vocabulary can ensure smooth and precise communication, such that people can convey ideas and enhance LSRW skills [100].
- **Layout design**: Effectiveness of English learning and design of the interface are significantly related because the interface is used to communicate with apps [90]. Thus, this point is included in the benchmarking checklist.

N. K. Ibrahim *et al.*: Multi-Criteria Evaluation and Benchmarking for Young Learners' English Language Mobile Applications

IEEE *Access*

TABLE 11. Comparison of scenarios and their related comparison points.

Scenario		Issue	Benchmark study1 [9]	Benchmark study2 [14]	Benchmark study3 [38]	Benchmark study4 [36]	Benchmark study5 [37]	Benchmark study6 [40]	Proposed DM study
First scenario	Content evaluation	Listening skill	✗	✗	✗	✗	✗	✗	✓
		Speaking skill	✗	✗	✓	✗	✗	✗	✓
		Reading skill	✗	✗	✗	✗	✗	✗	✓
		Writing skill	✗	✗	✗	✗	✗	✗	✓
		Vocabulary	✓	✗	✗	✗	✗	✓	✓
	Usability evaluation	Layout design	✓	✓	✓	✓	✓	✓	✗
		Motivation	✓	✓	✓	✓	✓	✗	✗
		Ease of use	✓	✓	✓	✓	✓	✓	✗
		Usefulness	✓	✗	✓	✓	✓	✓	✗
First scenario score			5 out of 9 issues	3 out of 9 issues	5 out of 9 issues	4 out of 9 issues	4 out of 9 issues	4 out of 9 issues	5 out of 9 issues
Second scenario		Multi evaluation criteria	✗	✗	✗	✗	✗	✗	✓
		Criteria weighting	✗	✗	✗	✗	✗	✗	✓
		Evaluation data variation	✗	✗	✗	✗	✗	✗	✓
		Multi criteria decision ranking	✗	✗	✗	✗	✗	✗	✓
Second scenario score			0 out of 4 issues	0 out of 4 issues	0 out of 4 issues	0 out of 4 issues	0 out of 4 issues	0 out of 4 issues	4 out of 4 issues
Third scenario		Validation	✗	✗	✗	✗	✗	✗	✓
		Evaluation	✗	✗	✗	✗	✗	✗	✓
Third scenario score			0 out of 2 issues	0 out of 2 issues	0 out of 2 issues	0 out of 2 issues	0 out of 2 issues	0 out of 2 issues	2 out of 2 issues
Total score			33.33%	20.00%	33.33%	26.66%	26.66%	26.66%	73.33%
Finding difference			40.00%	53.33%	40.00%	46.67%	46.67%	46.67%	

- **Motivation**: This point is a major factor in determining the success or failure in language learning. The aspects of motivation should be considered as one of the important elements related to English learning apps [91].
- **Ease of use**: Ease of use refers to the ease in learning and using apps [101]. Ease of use is essential in creating intention to use apps [102]. Thus, this point is included in this benchmarking checklist.
- **Usefulness**: This point reflects if an app usefulness evaluation has been provided. Usefulness can be defined as an insight into users' learning performance by using apps [103].
- **Multi-evaluation criteria**: This point displays whether the study has addressed multiple criteria during the app benchmarking process. Benchmarking is challenging because a decision is made on the basis of a set of attributes [81].
- **Criteria weighting**: This comparison point exhibits whether the criteria are assigned with the weights during the benchmarking process. One criterion may be

preferred more than others. The importance of each criterion in terms of the decision makers' preferences can be represented as weight [22], [81].
- **Evaluation data variation**: Multiple criteria generate a data variation is considered a multi-attribute decision problem [104] and must be handled accordingly [81]. Thus, this point is included in the benchmarking checklist.
- **Multi-criteria decision ranking**: This point indicates selecting the appropriate app after the evaluation process [81]. Therefore, this issue must be included in the benchmarking.
- **Validation**: Validation is a powerful way to reduce the likelihood of spreading false positive results [105]. Therefore, this issue is important and is included in the checklist.
- **Evaluation**: This point represents whether an evaluation has been provided and the proposed work is evaluated. Our evaluation aims to compare the performance of different studies on benchmarking and ranking English learning apps [106].

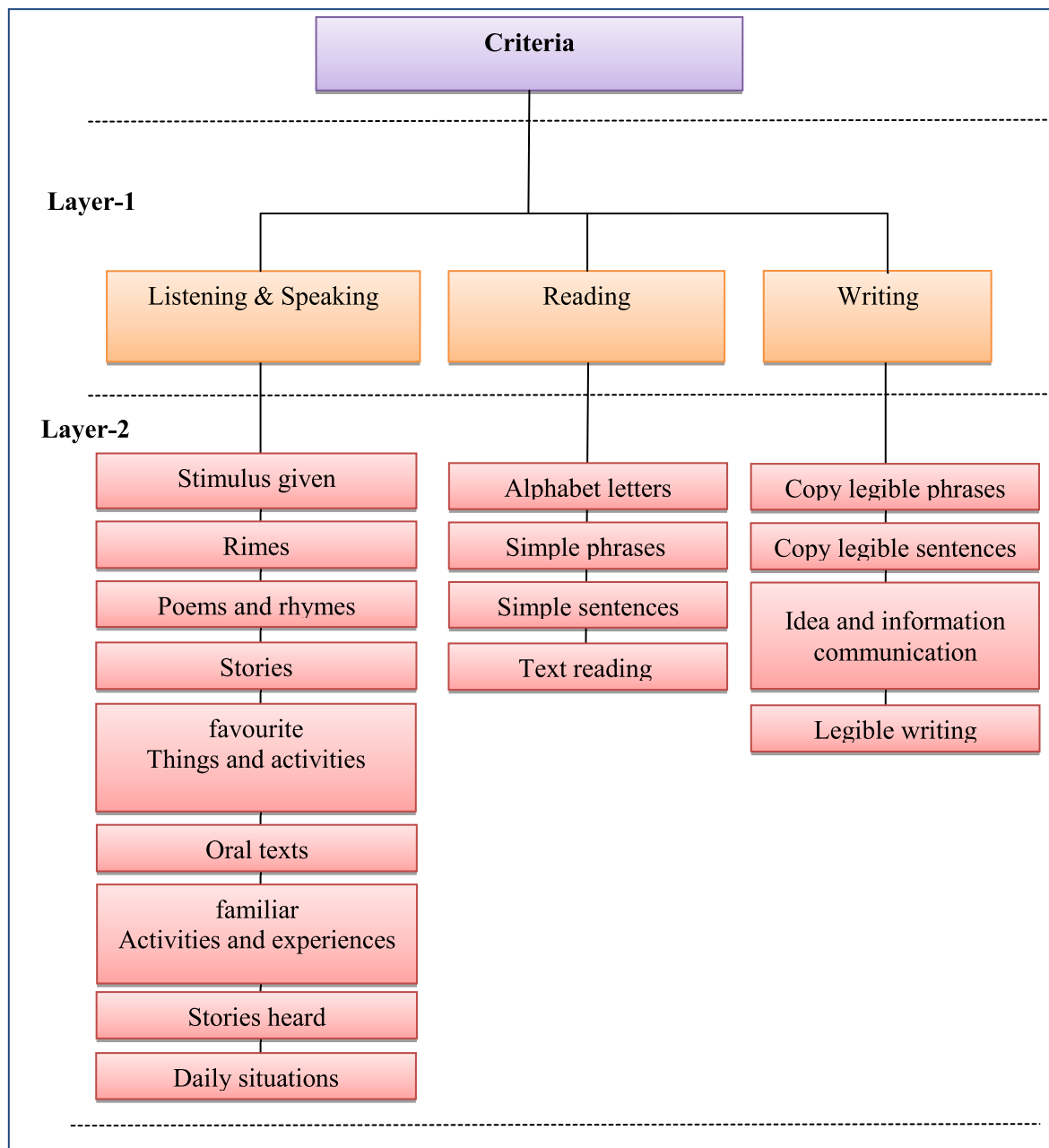**FIGURE 9.** Illustrates levels of evaluation criteria for evaluation the English learning apps based on KSPK 2016 standard.

TABLE 12. Comparison measurement scale.

Intensity of Importance	Definition
1	Equal importance
3	Moderately more important
5	Strongly more important
7	Very strongly more important
9	Extremely more important
2,4,6,8	Intermediate values

N. K. Ibrahim *et al.*: Multi-Criteria Evaluation and Benchmarking for Young Learners' English Language Mobile Applications

IEEE*Access*

TABLE 13. Comparison to determine the preference of most and least important criterion.

Main Criteria	Most Important	Least Important
Listening and Speaking		
Reading		
Writing		

TABLE 14. Comparison to determine the preference of most important criterion over other criteria.

Other Criteria / Most Important		

After defining the checklist comparison points, the comparison procedure is demonstrated. In those scenarios, 9, 4 and 2 out of 15 issues are highlighted for the first, second and last scenarios, respectively. Each comparison point within each scenario has gained 6.6667% from the overall performance (100 divided by 15 issues). Table 11 presents the checklist comparison between the proposed study and benchmark studies.

Table 11 shows that the benchmarking studies focus on the evaluation of layout design, motivation, ease of use and usefulness more than other issues. Only benchmark study 3 addresses the speaking skill issue, and benchmark studies 1 and 6 address the vocabulary. These studies are conducted through an app evaluation process only without benchmarking, result validation and evaluation processes. The proposed DM addresses 11 out of 15 issues, namely, listening skill, speaking skill, reading skill, writing skill, vocabulary, multi-evaluation criteria, criteria weighting, evaluation data variation, multi-criteria decision ranking, validation and evaluation issues. Worthy to note, the current study is based on KSPK standard (see Identification phase), therefore the other four issues namely, layout design, motivation, ease of use and usefulness are not addressed.

The differences in comparison studies are based on the scenarios and related comparison points also explained in Table 11. This table shows that the proposed study and benchmark study 1 exhibit an advantage over the five other benchmark studies in the first scenario with a total performance of 55.55% (5 out of 9 issues). In the second and third scenarios, the proposed study exhibits an advantage over the six benchmark studies with a total performance of 100% (4 out of 4 issues and 2 out of 2 issues, respectively).

However, the proposed DM study has covered 11 out of 15 issues in all scenarios (with a total performance of 73.33?%), whereas benchmark studies 1, 2, 3, 4, 5 and 6 covered 5, 3, 5, 4, 4 and 4 out of the 16 issues in all scenarios (with total performance of 33.33%, 20.00%, 33.33%, 26.66%, 26.66% and 26.66%, respectively).

The advantages and strengths of the issues that have been considered by the proposed DM and ignored by the benchmark studies are as follows:

- **Listening skill**: The first language skill we frequently learn is listening because children start listening and responding well to language even before talking [84]. Therefore, this issue is crucial in evaluating English learning apps.
- **Speaking skill**: Speaking must be mastered by students to be good communicators and to speak English fluently and accurately [107]. Thus, each English app should focus on this issue.
- **Reading skill**: Reading is important for learners to interact with written texts [108]. Through the ability to read well, learners can understand texts [109]. Consequently, English apps should consider this issue.
- **Writing skill**: Writing skill is typically considered a clear indication of whether or not learners considerably learnt English [110], [111]. Thus, this issue cannot be ignored in evaluating E-apps.
- **Vocabulary**: Vocabulary is central to English learning because students cannot understand others or express their ideas without adequate vocabulary [112]. Therefore, E-apps must address the vocabulary issue.
- **Multi-evaluation criteria**: Multi-criteria ranking is critical [81] for benchmarking English apps because this ranking is a complex decision-making problem based on multiple criteria. All criteria should be considered for assessment [19].
- **Criteria weighting**: Weighting technique plays an important role in benchmarking because the technique specifies the importance of the availability of each criterion against other criteria on the basis of expert judgement [81].
- **Evaluation data variation**: Handling data variation is important because it simplifies the selection decision with massive data [57].

IEEE Access

N. K. Ibrahim *et al.*: Multi-Criteria Evaluation and Benchmarking for Young Learners' English Language Mobile Applications

TABLE 15. Comparison to determine the preference of least important criterion over other criteria.

Criteria ╲	Least Important	

TABLE 16. Comparison to determine the most and least important criteria in level 2 of criteria (A).

sub-criteria A (Level 2)	Most Important	Least Important
Stimulus given		
Rimes		
Poems and rhymes		
Stories		
Favourite things and activities		
Oral texts		
Familiar activities and experiences		
Stories heard		
Daily situations		

TABLE 17. Comparison to determine the preference of most important criterion over the other criteria in level 2 of criteria (A).

Criteria ╲ Most Important									

- **Multi-criteria decision ranking**: This point indicates selecting the proper English learning app after the evaluation process [81].
- **Validation**: Benchmarking and ranking English learning apps are significant for learners. Thus, the validity of the selected procedure must be determined [105].
- **Evaluation**: The most relevant studies are compared, and the differences amongst them are determined [106], [113].

In summary, the statistical results for the evaluation process illustrate that the proposed DM exhibits an advantage over the six benchmark studies by 40.00%, 53.33%, 40.00%, 46.67%, 46.67% and 46.67%.

VI. CONCLUSION

The main contribution of this article is a DM for evaluating and benchmarking English learning apps for learners who are 5+ years of age. Six English learning apps were evaluated and ranked in terms of LSRW criteria and their sub-criteria, which were identified from the KSPK standard. In addition, the proposed DM facilitated the process of benchmarking these apps to help learners select suitable and reliable apps. The findings of this study emphasised three open issues of the evaluation criteria, namely, issues for multi-evaluation criteria, criterion importance and data variation. This study used integrated MCDM techniques that were regarded as solutions. The BWM technique was initially utilised to

N. K. Ibrahim *et al.*: Multi-Criteria Evaluation and Benchmarking for Young Learners' English Language Mobile Applications

IEEE*Access*

TABLE 18. Comparison to determine the preference of all criteria over the least important criterion in level 2 of criteria (A).

Criteria / Least Important	

TABLE 19. Comparison to determine the most and least important criteria in level 2 of criteria (B).

sub-criteria B (Level 2)	Most Important	Least Important
Alphabet letters		
Simple phrases		
Simple sentences		
Text reading		

TABLE 20. Comparison to determine the preference of most important criterion over the other criteria in level 2 of criteria (B).

Criteria / Most Important			

TABLE 21. Comparison to determine the preference of all criteria over the least important criterion in level 2 of criteria (B).

Criteria / Least Important	

assign weights for the identified criteria. Subsequently, internal and external TOPSIS techniques were used to benchmark and rank English learning apps. The results were then objectively validated. The statistical results indicated that the ranking results of English learning apps underwent a systematic ranking on the basis of internal and external TOPSIS aggregation. Finally, three main scenarios and a benchmarking checklist were provided for evaluation to demonstrate the performance of the proposed DM over the six other studies.

IEEE Access

N. K. Ibrahim *et al.*: Multi-Criteria Evaluation and Benchmarking for Young Learners' English Language Mobile Applications

TABLE 22. Comparison to determine the most and least important criteria in level 2 of criteria (C).

sub-criteria C (Level 2)	Most Important	Least Important
Copy legible phrases		
Copy legible sentences		
Idea and information communication		
Legible writing		

TABLE 23. Comparison to determine the preference of most important criterion over the other criteria in level 2 of criteria (C).

Most Important \ Criteria			

TABLE 24. Comparison to determine the preference of all criteria over the least important criterion in level 2 of criteria (C).

Criteria \ Least Important	

APPENDIX A
PAIRWISE COMPARISONS STANDARD
Section 1:
Dear Dr,

The aim behind this standard is to preferences comparison between criteria of evaluation and benchmarking English learning mobile applications for specifying the importance for each of which against others. This standard is a part of the research activities towards Master degree for Nu'as Kawther Ibrahim, a student at Universiti Pendidikan Sultan Idris (UPSI)/Malaysia.

Background:
Name:

Years of experience:

E-Mail:

Position:

Prior to answering the questions, it is important to understand the criteria assessed in arriving to a decision.

Many criteria used in this research for evaluation the English learning mobile apps for childhood. These evaluation criteria were divided into three main groups, namely, (1) Listening and Speaking, (2) Reading, and (3) Writing;

The Listening and Speaking group includes nine metrics of criteria, namely: (stimulus given, rimes, poems and rhymes, stories, favourite things and activities, oral texts, familiar activities and experiences, stories heard, and daily situations). Reading has four metrics (alphabet letters, simple phrases, simple sentences, and text reading), and Writing has four metrics (copy legible phrases, copy legible sentences, idea and information communication, legible writing). The following figure 9 illustrates the levels:

Section 2: comparison questions
Comparison measurement scale
The comparisons (relative importance) of each criterion are measured according to a numerical scale from 1 to 9. These relative scales (1 to 9), as shown in Table 12, Please use this scale in comparison.

1. **Main Criteria**

    A. **Listening and Speaking:** this criterion is used to evaluate the app with respect to Listening and Speaking skills.

N. K. Ibrahim *et al.*: Multi-Criteria Evaluation and Benchmarking for Young Learners' English Language Mobile Applications

IEEE *Access*

TABLE 25. Checklist form.

Criteria For Skills:	E-learning suitability	Lingokids	Fun English	FunWith-Flupe	First Words	Montessori	Spelling Bee
listening and speaking criteria							
Stimulus given: Are pupils able to listen to and respond to stimulus given (environmental sounds, voice sounds, rhythm and rhyme, and alliteration)?							
Rimes: Are pupils able to listen to and identify rimes in nursery rhymes and songs?							
Poems and rhymes: Are pupils able to listen to and recite poems and rhymes?							
Stories: Are pupils able to listen to and respond to stories?							
Favourite things and activities: Are pupils able to talk about favourite things and activities?							
Oral texts: Are pupils able to listen to and respond to oral texts?							
Familiar activities and experiences: Are pupils able to talk about familiar activities and experiences?							
Stories heard: Are pupils able to talk about stories heard?							
Daily situations: Are pupils able to role play familiar daily situations?							
Reading criteria							
Alphabet letters: Are pupils able to recognise and sound out letters of the alphabet?							
Sounds in a word: Are pupils able to recognise and sound out initial, medial and ending sounds in a word?							
Blend sounds: Are pupils able to blend phonemes (sounds) to form single syllable words?							
Frequency/sight words: Are pupils able to recognise and read high frequency/sight words?							
Simple phrases: Are pupils able to read simple phrases?							
Simple sentences: Are pupils able to read simple sentences?							
Independency: Are pupils able to read texts independently?							
Text reading: Are pupils able to read and respond to texts read?							
Writing criteria							
Copy legible phrases: Are pupils able to copy simple phrases in legible print?							
Copy legible sentences: Are pupils able to copy simple sentences in legible print?							
Idea and information communication: Are pupils able to communicate ideas and information by using drawing, marks, symbols and writing with invented spelling?							
Legible writing: Are pupils able to write words and phrases in legible print?							

IEEE Access

N. K. Ibrahim *et al.*: Multi-Criteria Evaluation and Benchmarking for Young Learners' English Language Mobile Applications

TABLE 26. Results obtained from the second expert.

Main Criteria Sub Criteria Alternatives (apps)	Listening and Speaking									Reading				Writing			
	Stimulus given (sg)	Rimes (ri)	Poems and rhymes (p&r)	Stories (st)	favourite things and activities (ft&a)	Oral texts (ot)	Familiar activities and experiences (fa&e)	Stories heard (sh)	Daily situations (ds)	Alphabet letters (al)	Simple phrases (sp)	Simple sentences (ss)	Text reading (tr)	Copy legible phrases (clp)	Copy legible sentences (cls)	Idea and information communication (i&ic)	Legible writing (lw)
Lingokids	1	1	1	1	1	1	1	0	1	1	0	0	1	1	1	1	1
Fun English	1	0	0	0	1	1	1	0	1	1	0	1	1	1	0	1	1
FunWithFlupe	1	0	0	0	1	0	1	0	0	1	1	1	1	1	1	1	1
First Words	1	0	0	0	1	1	1	0	1	1	0	0	1	1	0	1	1
Montessori	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Spelling Bee	1	0	0	0	1	1	1	0	1	1	1	0	1	1	1	1	1

TABLE 27. Results obtained from the third expert.

Main Criteria Sub Criteria Alternatives (apps)	Listening and Speaking									Reading				Writing			
	Stimulus given (sg)	Rimes (ri)	Poems and rhymes (p&r)	Stories (st)	favourite things and activities (ft&a)	Oral texts (ot)	Familiar activities and experiences (fa&e)	Stories heard (sh)	Daily situations (ds)	Alphabet letters (al)	Simple phrases (sp)	Simple sentences (ss)	Text reading (tr)	Copy legible phrases (clp)	Copy legible sentences (cls)	Idea and information communication (i&ic)	Legible writing (lw)
Lingokids	1	1	1	1	1	1	1	0	1	1	0	0	0	1	1	1	0
Fun English	1	1	1	0	1	0	1	0	0	1	0	1	1	1	1	1	1
FunWithFlupe	1	0	0	0	1	1	1	0	0	1	1	1	1	1	1	1	1
First Words	1	0	0	1	1	0	1	0	1	1	1	0	1	1	0	1	1
Montessori	1	0	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1
Spelling Bee	1	0	0	1	0	1	1	0	0	1	1	0	1	0	1	0	1

B. Reading : this criterion is used to evaluate the app with respect to Reading skill.

C. Writing : this criterion is used to evaluate the app with respect to Writing skill.

Questions

1.1. Could you indicate, which of these three criteria you consider the MOST important and which one you consider the LEAST important by marking the box? Please in table 13, marking the cell of in front of the MOST important criterion and marking the cell of in front of the LEAST important criterion.

You have selected X criterion as the MOST important criterion and Y criterion as the LEAST important criterion

1.2. Please determine your preference of this criterion (X) over the other criteria by using 1 to 9 measurement scale.

Please write the X criterion that you selected as the most important criteria in green cell and the other

N. K. Ibrahim et al.: Multi-Criteria Evaluation and Benchmarking for Young Learners' English Language Mobile Applications

IEEE Access

TABLE 28. Results obtained from the fourth expert.

Main Criteria	Listening and Speaking									Reading				Writing			
Sub Criteria	Stimulus given (sg)	Rimes (ri)	Poems and rhymes (p&r)	Stories (st)	favourite things and activities (ft&a)	Oral texts (ot)	Familiar activities and experiences (fa&e)	Stories heard (sh)	Daily situations (ds)	Alphabet letters (al)	Simple phrases (sp)	Simple sentences (ss)	Text reading (tr)	Copy legible phrases (clp)	Copy legible sentences (cls)	Idea and information communication (i&ic)	Legible writing (lw)
Alternatives (apps)																	
Lingokids	1	1	0	0	0	0	1	0	0	0	1	1	0	0	1	1	1
Fun English	1	1	1	0	1	1	0	0	1	1	1	0	1	1	0	1	1
FunWithFlupe	1	0	1	1	0	0	1	0	0	0	1	1	1	0	1	1	1
First Words	1	1	0	0	1	1	1	0	0	0	0	1	0	1	0	1	1
Montessori	1	0	1	1	1	0	0	1	1	1	1	1	1	0	1	1	1
Spelling Bee	1	0	1	1	1	1	1	0	1	1	1	0	1	1	0	1	1

TABLE 29. Results obtained from the fifth expert.

Main Criteria	Listening and Speaking									Reading				Writing			
Sub Criteria	Stimulus given (sg)	Rimes (ri)	Poems and rhymes (p&r)	Stories (st)	favourite things and activities (ft&a)	Oral texts (ot)	Familiar activities and experiences (fa&e)	Stories heard (sh)	Daily situations (ds)	Alphabet letters (al)	Simple phrases (sp)	Simple sentences (ss)	Text reading (tr)	Copy legible phrases (clp)	Copy legible sentences (cls)	Idea and information communication (i&ic)	Legible writing (lw)
Alternatives (apps)																	
Lingokids	0	1	1	1	0	0	1	1	0	0	0	0	1	1	1	1	1
Fun English	1	1	1	1	1	1	0	1	1	0	0	1	0	0	0	0	0
FunWithFlupe	0	0	1	1	0	0	0	0	0	1	1	0	1	1	1	1	1
First Words	1	0	0	0	1	1	1	0	1	1	0	0	0	0	0	0	1
Montessori	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1	1	1
Spelling Bee	1	0	0	0	0	0	1	0	1	1	1	0	1	1	1	1	1

criteria in the grey cell in table (14), and then write your preferences value.

1.3. You have selected Y criterion as the LEAST important criterion.

Please determine your preference of all criteria over the Y criterion that you selected as LEAST important criterion by using 1 to 9 measurement scale.

Please write the Y criterion that you selected as LEAST important criteria in green cell and the other criteria

in the grey cells in table (15), and then write your preferences value.

2. The sub-criteria (A) - (level 2)

A. **Stimulus given:** listen to and respond to stimulus given (environmental sounds, voice sounds, rhythm and rhyme, and alliteration).

B. **Rimes:** listen to and identify rimes in nursery rhymes and songs.

IEEE Access

N. K. Ibrahim *et al.*: Multi-Criteria Evaluation and Benchmarking for Young Learners' English Language Mobile Applications

TABLE 30. Results obtained from the sixth expert.

Main Criteria / Sub Criteria / Alternatives (apps)	Listening and Speaking									Reading				Writing			
	Stimulus given (sg)	Rimes (ri)	Poems and rhymes (p&r)	Stories (st)	favourite things and activities (ft&a)	Oral texts (ot)	Familiar activities and experiences (fa&e)	Stories heard (sh)	Daily situations (ds)	Alphabet letters (al)	Simple phrases (sp)	Simple sentences (ss)	Text reading (tr)	Copy legible phrases (clp)	Copy legible sentences (cls)	Idea and information communication (i&ic)	Legible writing (lw)
Lingokids	0	1	1	1	1	0	0	1	0	0	1	1	0	0	1	1	0
Fun English	1	1	1	0	1	0	0	1	1	0	1	1	0	0	0	0	0
FunWithFlupe	1	0	1	1	0	1	1	0	0	1	1	1	0	1	1	1	1
First Words	0	1	0	1	0	1	1	1	0	1	0	0	1	1	0	1	0
Montessori	1	0	1	1	1	1	1	1	1	0	1	0	1	1	0	0	1
Spelling Bee	1	1	1	1	0	1	1	0	1	0	1	0	1	0	1	0	1

C. **Poems and rhymes**: listen to and recite poems and rhymes.

D. **Stories:** Listen to and respond to stories.

E. **Favourite things and activities:** talk about favourite things and activities.

F. **Oral texts:** listen to and respond to oral texts.

G. **Familiar activities and experiences:** talk about familiar activities and experiences.

H. **Stories heard:** talk about stories heard.

I. **Daily situations:** role play familiar daily situations.

Questions

2.1. Could you indicate which one of these criteria (**sub-criteria A (Level 2)**) consider the MOST important and which one you find the LEAST important? Please in table 16, marking the cell of in front of the MOST important criterion and marking the cell of in front of the LEAST important criterion.
You have selected X criterion as the MOST important criterion and Y criterion as the LEAST important criterion

2.2. Please determine your preference of the criterion (X) over the other criteria by using 1 to 9 measurement scale.
Please write the X criterion that you selected as most important criterion in green cell and the other criteria in the grey cells in table (17), and then write your preferences value.

2.3. You have selected Y criterion as the LEAST important criterion.

Please determine your preference of all criteria over the Y criteria that you selected as LEAST important criterion by using 1 to 9 measurement scale.
Please write the Y criterion that you selected as LEAST important criteria in green cell and the other criteria in the grey cells in table (18), and then write your preferences value.

3. **The sub-criteria (B) - (level 2)**

a) **Alphabet letters:** recognise and sound out letters of the alphabet.

b) **Simple phrases:** read simple phrases.

c) **Simple sentences:** read simple sentences.

d) **Text reading:** read and respond to texts read.

Questions

3.1. Could you indicate which one of these criteria (**sub-criteria B (Level 2)**) consider the MOST important and which one you find the LEAST important? Please in table 19, marking the cell of in front of the MOST important criterion and marking the cell of in front of the LEAST important criterion.
You have selected X criterion as the MOST important criterion and Y criterion as the LEAST important criterion

3.2. Please determine your preference of the criterion (X) over the other criteria by using 1 to 9 measurement scale.
Please write the X criterion that you selected as most important criterion in green cell and the other criteria in the grey cells in table (20), and then write your preferences value.

TABLE 31. The results of the BWM method for weight preferences of the criteria of English learning apps evaluation (second and third experts).

Expert 2							
Level 1 of Criteria: Main Criteria							
List of criteria	Best criterion	Other Criteria	Scores	Other criteria	Worst criterion	Scores	Weight
Listening and speaking	Writing	Listening and speaking	2	Listening and speaking	reading	4	0.325
Reading		Reading	5	writing		5	0.100
Writing		....	....	....		....	0.575
Consistency: 0.033							
Level 2 of Criteria: sub criteria of Listening and speaking							
List of criteria	Best criterion	Other Criteria	Scores	Other criteria	Worst criterion	Scores	Weight
Stimulus given	Favourite things and activities	Stimulus given	5	Stimulus given	Rimes	5	0.062
Rimes		Rimes	8	Poems and rhymes		4	0.024
Poems and rhymes		Poems and rhymes	6	Stories		4	0.051
Stories		Stories	2	Favourite things and activities		8	0.154
Favourite things and activities		Oral texts	4	Oral texts		3	0.250
Oral texts		Familiar activities and experiences	2	Familiar activities and experiences		5	0.048
Familiar activities and experiences		Stories heard	3	Stories heard		4	0.154
Stories heard		Daily situations	2	Daily situations		5	0.103
Daily situations		....	....	....		....	0.154
Consistency: 0.013							
Level 3 of Criteria: sub criteria of Reading							
List of criteria	Best criterion	Other Criteria	Scores	Other criteria	Worst criterion	Scores	Weight
Alphabet letters	Alphabet letters	Simple phrases	4	Alphabet letters	Text reading	9	0.564
Simple phrases		Simple sentences	3	Simple phrases		5	0.164
Simple sentences				Simple sentences		5	0.220
Text reading		Text reading	9				0.052
Consistency: 0.018							
Level 4 of Criteria: sub criteria of Writing							
List of criteria	Best criterion	Other Criteria	Scores	Other criteria	Worst criterion	Scores	Weight
Copy legible phrases	Idea and information communication	Copy legible phrases	2	Copy legible phrases	Legible writing	2	0.254
Copy legible sentences		Copy legible sentences	3	Copy legible sentences		4	0.193
Idea and information communication		Legible writing	5	Idea and information communication		5	0.478
Legible writing		....	....	....		....	0.075
Consistency: 0.045							
Expert 3							
Level 1 of Criteria: Main Criteria							
List of criteria	Best criterion	Other Criteria	Scores	Other criteria	Worst criterion	Scores	Weight
Listening and speaking	Listening and speaking	reading	3	Listening and speaking	reading	3	0.542
Reading		writing	2	writing		2	0.166
Writing		....	....	....		....	0.292
Consistency: 0.042							
Level 2 of Criteria: sub criteria of Listening and speaking							
List of criteria	Best criterion	Other Criteria	Scores	Other criteria	Worst criterion	Scores	Weight
Stimulus given	Stories	Stimulus given	4	Stimulus given	Poems and rhymes	3	0.070
Rimes		Rimes	3	Rimes		4	0.095
Poems and rhymes		Poems and rhymes	5	Stories		5	0.036
Stories		Favourite things and activities	2	Favourite things and activities		3	0.230

IEEE Access

N. K. Ibrahim *et al.*: Multi-Criteria Evaluation and Benchmarking for Young Learners' English Language Mobile Applications

TABLE 31. *(Continued.)* The results of the BWM method for weight preferences of the criteria of English learning apps evaluation (second and third experts).

Favourite things and activities		Oral texts	2	Oral texts		2	0.140
Oral texts		Familiar activities and experiences	4	Familiar activities and experiences		2	0.122
Familiar activities and experiences		Stories heard	2	Stories heard		2	0.045
Stories heard		Daily situations	2	Daily situations		3	0.122
Daily situations		....	....	....		....	0.140
Consistency: 0.022							
Level 3 of Criteria: sub criteria of Reading							
List of criteria	Best criterion	Other Criteria	Scores	Other criteria	Worst criterion	Scores	Weight
Alphabet letters		Simple phrases	8	Alphabet letters		8	0.518
Simple phrases	Alphabet letters	Simple sentences	2	Simple sentences	Simple phrases	3	0.069
Simple sentences							0.241
Text reading		Text reading	3	Text reading		2	0.172
Consistency: 0.008							
Level 4 of Criteria: sub criteria of Writing							
List of criteria	Best criterion	Other Criteria	Scores	Other criteria	Worst criterion	Scores	Weight
Copy legible phrases		Copy legible phrases	8	Copy legible phrases		7	0.117
Copy legible sentences	Idea and information communication	Copy legible sentences	7	Copy legible sentences	Legible writing	6	0.134
Idea and information communication		Legible writing	9	Idea and information communication		9	0.698
Legible writing		....	....	....		....	0.051
Consistency: 0.046							

3.3. You have selected Y criterion as the LEAST important criterion.

Please determine your preference of all criteria over the Y criteria that you selected as LEAST important criterion by using 1 to 9 measurement scale.

Please write the Y criterion that you selected as LEAST important criterion in green cell and the other criteria in the grey cells in table (21), and then write your preferences value.

4. The sub-criteria (C) - (level 2)

A. **Copy legible phrases:** copy simple phrases in legible print.

B. **Copy legible sentences:** copy simple sentences in legible print.

C. **Idea and information communication:** communicate ideas and information by using drawing, marks, symbols and writing with invented spelling.

D. **Legible writing:** write words and phrases in legible print.

Questions

4.1. Could you indicate which one of these criteria (**sub-criteria C (Level 2)**) consider the MOST important and which one you find the LEAST important? Please in table 22, marking the cell of in front of the MOST important criterion and marking the cell of in front of the LEAST important criterion.

You have selected X criterion as the MOST important criterion and Y criterion as the LEAST important criterion

4.2 Please determine your preference of the criterion (X) over the other criteria by using 1 to 9 measurement scale.

Please write the X criterion that you selected as most important criterion in green cell and the other criteria in the grey cells in table (23), and then write your preferences value.

4.3. You have selected Y criterion as the LEAST important criterion.

Please determine your preference of all criteria over the Y criteria that you selected as LEAST important criterion by using 1 to 9 measurement scale.

Please write the Y criterion that you selected as LEAST important criterion in green cell and the other criteria in the grey cells in table (24), and then write your preferences value.

In case you have any inquiry or wish to know the result, please contact:

Email: nuas.almansori@gmail.com **Mobile phone:** 00601161344406

······· **Thank you for Your Time** ·······

N. K. Ibrahim *et al.*: Multi-Criteria Evaluation and Benchmarking for Young Learners' English Language Mobile Applications

IEEE *Access*

TABLE 32. Ranking results based on the second and third experts' weights.

NO.	App	S-	S*	Rank Value	Final Rank
	Second experts				
1-	Lingokids	0.0814	0.0343	0.7033	2
2-	Fun English	0.0393	0.0778	0.3355	6
3-	FunWithFlupe	0.0795	0.0482	0.6228	3
4-	First Words	0.0403	0.0745	0.3510	5
5-	Montessori	0.0796	0.0270	0.7464	1
6-	Spelling Bee	0.0604	0.0637	0.4870	4
	Third experts				
1)	Lingokids	0.0844	0.0537	0.6109	2
2)	Fun English	0.0610	0.0892	0.4059	3
3)	FunWithFlupe	0.0563	0.0941	0.3744	5
4)	First Words	0.0504	0.0859	0.3696	6
5)	Montessori	0.1100	0.0302	0.7845	1
6)	Spelling Bee	0.0579	0.0906	0.3899	4

APPENDIX B
ENGLISH LEARNING APPS EVALUATION CHECKLIST FORM

University Pendidikan Sultan Idris
Faculty of Art, Computing and Creative Industry

MULTI-CRITERIA ANALYSIS FOR EVALUATION AND BENCHMARKING YOUNG LEARNERS ENGLISH LANGUAGE MOBILE APPLICATIONS IN TERMS OF LSRW SKILLS BASED KSPK STANDARD

Dear valued Dr,

This ckecklist form is designed for the purpose of a study that will help me (Nu'as Kawther Ibrahim) to complete a research as a requirement for Master degree at Universiti Pendidikan Sultan Idris (UPSI)/Malaysia. It is designed to evaluate six English learning mobile apps for early childhood (at age 5+) with respect to the four main language skills (Listening, Speaking, Reading, and Writing). As you are the best one to give the correct picture of your experience in English learning for early childhood, please respond to the following questions frankly and honestly.

Background:
Name:
Years of experience:
E-Mail:
Position:

In the box of your answer, please, place a "✔" mark if you agree, or "✗" mark if you do not.

APPENDIX C
RESULTS
See Tables 26–32.

REFERENCES

[1] Y.-T. C. Yang and C.-Y. Chan, "Comprehensive evaluation criteria for English learning websites using expert validity surveys," *Comput. Educ.*, vol. 51, no. 1, pp. 403–422, 2008.
[2] M. Robson, "The English effect: The impact of English, what it's worth to the UK and why it matters to the world," *British Council*, pp. 1–24, 2013.
[3] F. Walter and F. Wrester, "Early childhood development: The key to a full and productive life," *J. Childhood Develop.*, vol. 23, no. 232, p. 23, 2009.
[4] Y. Chen, D. Zhou, Y. Wang, and J. Yu, "Application of augmented reality for early childhood English teaching," in *Proc. Int. Symp. Educ. Technol. (ISET)*, Jun. 2017, pp. 111–115.
[5] R. Gangaiamaran and M. Pasupathi, "Review on use of mobile apps for language learning," *J. Appl. Eng. Res.*, vol. 12, no. 21, pp. 11242–11251, 2017.
[6] G. Oakley and U. Imtinan, "Supporting children's literacy learning in low- and middle-income countries through m-learning," in *Mobile Technologies in Children's Language and Literacy: Innovative Pedagogy in Preschool and Primary Education*. Bingley, U.K.: Emerald Publishing, 2018, pp. 155–175.
[7] K. O. Jeong, "University students' perception and motivation of using digital applications as effective english learning tools," in *Proc. Int. Conf. Platform Technol. Service (PlatCon)*, Feb. 2017, pp. 3–6.
[8] Y. Kim and D. Smith, "Pedagogical and technological augmentation of mobile learning for young children interactive learning environments," *Interact. Learn. Environ.*, vol. 25, no. 1, pp. 4–16, Nov. 2015.
[9] M. F. Kamaruzaman and I. H. Zainol, "Behavior response among secondary school students development towards mobile learning application," in *Proc. IEEE Colloq. Humanities Sci. Eng. (CHUSER)*, Dec. 2012, pp. 589–592.
[10] B. S. Gomathi, "Enriching the skills of rural students with effective methods of teaching English language using LSRW Skills," *Int. J. Educ. Inf. Stud.*, vol. 4, no. 2, pp. 65–69, 2014.
[11] X. L. Pham, T. H. Nguyen, W. Y. Hwang, and G. D. Chen, "Effects of push notifications on learner engagement in a mobile learning app," in *Proc. IEEE 16th Int. Conf. Adv. Learn. Technol. (ICALT)*, Jul. 2016, pp. 90–94.
[12] M. M. Elaish, L. Shuib, N. A. Ghani, E. Yadegaridehkordi, and M. Alaa, "Mobile learning for English language acquisition: Taxonomy, challenges, and recommendations," *IEEE Access*, vol. 5, pp. 19033–19047, 2017.

IEEE *Access*

N. K. Ibrahim *et al.*: Multi-Criteria Evaluation and Benchmarking for Young Learners' English Language Mobile Applications

[13] R. B. Sadiq, N. Cavus, and D. Ibrahim, "Mobile application based on CCI standards to help children learn English as a foreign language," *Interact. Learn. Environ.*, to be published.

[14] R.-A. Pilar, A. Jorge, and C. Cristina, "The use of current mobile learning applications in EFL," *Procedia-Soc. Behav. Sci.*, vol. 103, pp. 1189–1196, Nov. 2013.

[15] M. Kalz, N. Lenssen, M. Felzen, R. Rossaint, B. Tabuenca, M. Specht, and M. Skorning, "Smartphone apps for cardiopulmonary resuscitation training and real incident support: A mixed-methods evaluation study," *J. Med. Internet Res.*, vol. 16, no. 3, p. e89, 2014.

[16] S. Basaran and O. J. Aduradola, "A multi-criteria decision making to rank Android based mobile applications for mathematics," *Int. J. Adv. Comput. Sci. Appl.*, vol. 9, no. 7, pp. 99–107, 2018.

[17] Q. M. Yas, A. A. Zaidan, B. B. Zaidan, B. Rahmatullah, and H. A. Karim, "Comprehensive insights into evaluation and benchmarking of real-time skin detectors: Review, open issues & challenges, and recommended solutions," *Measurement*, vol. 114, pp. 243–260, Jan. 2018.

[18] M. Leonid, "Transforming computer lab into a mini HPC cluster," Tech. Rep., 2013, pp. 252–255.

[19] A. R. Karimi, N. Mehrdadi, S. J. Hashemian, G. R. N. Bidhendi, and R. T. Moghaddam, "Selection of wastewater treatment process based on the analytical hierarchy process and fuzzy analytical hierarchy process methods," *Int. J. Environ. Sci. Technol.*, vol. 8, no. 2, pp. 267–280, 2011.

[20] A. A. Zaidan, B. B. Zaidan, M. Hussain, A. Haiqi, M. L. M. Kiah, and M. Abdulnabi, "Multi-criteria analysis for OS-EMR software selection problem: A comparative study," *Decis. Support Syst.*, vol. 78, pp. 15–27, Oct. 2015.

[21] B. N. Abdullateef, N. F. Elias, H. Mohamed, A. A. Zaidan, and B. B. Zaidan, "An evaluation and selection problems of OSS-LMS packages," *SpringerPlus*, vol. 5, no. 1, p. 248, Dec. 2016.

[22] N. I. Jaini, "An efficient ranking analysis in multi-criteria decision making," School Mech., Aerosp. Civil Eng., Univ. Manchester, Manchester, U.K., 2017.

[23] B. B. Zaidan, A. A. Zaidan, H. A. Karim, and N. N. Ahmad, "A new digital watermarking evaluation and benchmarking methodology using an external group of evaluators and multi-criteria analysis based on 'large-scale data,'" *Softw., Pract. Exper.*, vol. 47, no. 10, pp. 1365–1392, 2017.

[24] B. B. Zaidan and A. A. Zaidan, "Software and hardware FPGA-based digital watermarking and steganography approaches: Toward new methodology for evaluation and benchmarking using multi-criteria decision-making techniques," *J. Circuits, Syst. Comput.*, vol. 26, no. 7, 2017, Art. no. 1750116.

[25] Z. Chen and W. Yang, "A new multiple attribute group decision making method in intuitionistic fuzzy setting," *Appl. Math. Model.*, vol. 35, no. 9, pp. 4424–4437, 2011.

[26] B. B. Zaidan, A. A. Zaidan, H. A. Karim, and N. N. Ahmad, "A new approach based on multi-dimensional evaluation and benchmarking for data hiding techniques," *Int. J. Inf. Technol. Decis. Making*, to be published.

[27] B. Rahmatullah, A. A. Zaidan, F. Mohamed, and A. Sali, "Multi-complex attributes analysis for optimum GPS baseband receiver tracking channels selection," in *Proc. 4th Int. Conf. Control, Decis. Inf. Technol. (CoDIT)*, Apr. 2017, pp. 1084–1088.

[28] A. Roozbahani, B. Zahraie, and M. Tabesh, "PROMETHEE with precedence order in the criteria (PPOC) as a new group decision making aid: An application in urban water supply management," *Water Resour. Manage.*, vol. 26, no. 12, pp. 3581–3599, 2012.

[29] J. Liu, P. Liu, S.-F. Liu, X.-Z. Zhou, and T. Zhang, "A study of decision process in MCDM problems with large number of criteria," *Int. Trans. Oper. Res.*, vol. 22, no. 2, pp. 237–264, 2015.

[30] K. Mohammed, A. A. Zaidan, B. B. Zaidan, O. S. Albahri, M. A. Alsalem, A. S. Albahri, A. Hadi, and M. Hashim, "Real-time remote-health monitoring systems: A review on patients prioritisation for multiple-chronic diseases, taxonomy analysis, concerns and solution procedure," *J. Med. Syst.*, vol. 43, no. 7, p. 223, 2019.

[31] V. Podvezko, "Comparative analysis of MCDA methods SAW and COPRAS," *Eng. Econ.*, vol. 22, no. 2, pp. 134–146, 2011.

[32] X. Deng, P. Vroman, X. Zeng, and B. Laouisset, "Intelligent decision support tools for multicriteria product design," in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, Oct. 2010, pp. 1223–1230.

[33] M. Khatari, A. A. Zaidan, B. B. Zidan, O. S. Albahri, and M. A. Alsalem, "Multi-criteria evaluation and benchmarking for active queue management methods: Open issues, challenges and recommended pathway solutions," *Int. J. Inf. Technol. Decis. Making*, vol. 18, no. 4, pp. 1187–1242, 2019.

[34] A. A. Zaidan, B. B. Zaidan, M. A. Alsalem, O. S. Albahri, A. S. Albahri, and M. Y. Qahtan, "Multi-agent learning neural network and Bayesian model for real-time IoT skin detectors: A new evaluation and benchmarking methodology," *Neural Comput. Appl.*, to be published.

[35] C. K. Lim, K. L. Tan, A. A. Zaidan, and B. B. Zaidan, "A proposed methodology of bringing past life in digital cultural heritage through crowd simulation: A case study in George Town, Malaysia," *Multimedia Tools Appl.*, to be published.

[36] H. M. Al-Otaibi, R. A. Alamer, and H. S. Al-Khalifa, "The next generation of language labs: Can mobiles help? A case study," *Comput. Hum. Behav.*, vol. 59, pp. 342–349, Jun. 2016.

[37] Y. T. Chuang, "MEMIS: A mobile-supported English-medium instruction system," *Telematics Inform.*, vol. 34, no. 2, pp. 640–656, 2017.

[38] K. Segaran, A. Z. M. Ali, and T. W. Hoe, "Usability and user satisfaction of 3D talking-head mobile assisted language learning (MALL) app for non-native speakers," *Procedia-Soc. Behav. Sci.*, vol. 131, pp. 4–10, May 2014.

[39] Z. A. Silfiani and Y. S. Rezeki, "Teaching pronunciation in speaking using prosody pyramid," *J. Pendidikan Pembelajaran*, vol. 6, no. 10, pp. 1–10, 2017.

[40] L.-K. Lee, C.-H. Chau, C.-H. Chau, and C.-T. Ng, "Using augmented reality to teach kindergarten students English vocabulary," in *Proc. Int. Symp. Educ. Technol.*, Jun. 2017, pp. 53–57.

[41] Q. M. Yas, A. A. Zadain, B. B. Zaidan, M. B. Lakulu, and B. Rahmatullah, "Towards on develop a framework for the evaluation and benchmarking of skin detectors based on artificial intelligent models using multi-criteria decision-making techniques," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 31, no. 3, 2017, Art. no. 1759002.

[42] F. M. Jumaah, A. A. Zaidan, B. B. Zaidan, R. Bahbibi, M. Y. Qahtan, and A. Sali, "Technique for order performance by similarity to ideal solution for solving complex situations in multi-criteria optimization of the tracking channels of GPS baseband telecommunication receivers," *Telecommun. Syst.*, vol. 68, no. 3, pp. 425–443, Jul. 2018.

[43] M. M. Salih, B. B. Zaidan, A. A. Zaidan, and M. A. Ahmed, "Survey on fuzzy TOPSIS state-of-the-art between 2007–2017," *Comput. Oper. Res.*, vol. 104, pp. 207–227, Apr. 2019.

[44] N. Kalid, A. A. Zaidan, B. B. Zaidan, O. H. Salman, M. Hashim, and H. Muzammil, "Based real time remote health monitoring systems: A review on patients prioritization and related 'big data' using body sensors information and communication technology," *J. Med. Syst.*, vol. 42, no. 2, p. 30, 2018.

[45] F. M. Jumaah, A. A. Zadain, B. B. Zaidan, A. K. Hamzah, and R. Bahbibi, "Decision-making solution based multi-measurement design parameter for optimization of GPS receiver tracking channels in static and dynamic real-time positioning multipath environment," *Measurement*, vol. 118, pp. 83–95, Mar. 2018.

[46] S. Christodoulou, N. Karacapilidis, and M. Tzagarakis, "Advancing collaborative decision making through alternative visualizations and reasoning mechanisms," in *Intelligent Decision Technologies* (Frontiers in Artificial Intelligence and Applications), vol. 241, R. Neves-Silva, J. Watada, G. Phillips-Wren, L. C. Jain, and R. J. Howlett, Eds. 2013, pp. 38–47.

[47] O. H. Salman, A. A. Zaidan, B. B. Zaidan, Naserkalid, and M. Hashim, "Novel methodology for triage and prioritizing using 'big data' patients with chronic heart diseases through telemedicine environmental," *Int. J. Inf. Technol. Decis. Making*, vol. 16, no. 5, pp. 1211–1245, Sep. 2017.

[48] B. B. Zaidan and A. A. Zaidan, "Comparative study on the evaluation and benchmarking information hiding approaches based multi-measurement analysis using TOPSIS method with different normalisation, separation and context techniques," *Measurement*, vol. 117, pp. 277–294, May 2018.

[49] A. A. Zaidan, "A review on smartphone skin cancer diagnosis apps in evaluation and benchmarking: Coherent taxonomy, open issues and recommendation pathway solution," *Health Technol.*, vol. 8, no. 4, pp. 223–238, Sep. 2018.

[50] S. Christodoulou, N. Karacapilidis, and M. Tzagarakis, "Exploiting alternative knowledge visualizations and reasoning mechanisms to enhance collaborative decision making," in *Intelligent Decision Technology Support in Practice*. Cham, Switzerland: Springer, 2016, pp. 89–106.

N. K. Ibrahim *et al.*: Multi-Criteria Evaluation and Benchmarking for Young Learners' English Language Mobile Applications

IEEE Access

[51] M. A. Alsalem, A. A. Zaidan, B. B. Zaidan, M. Hashim, O. S. Albahri, A. S. Albahri, A. Hadi, and K. I. Mohammed, "Systematic review of an automated multiclass detection and classification system for acute Leukaemia in terms of evaluation and benchmarking, open challenges, issues and methodological aspects," *J. Med. Syst.*, vol. 42, no. 11, p. 204, 2018.

[52] A. A. Zaidan, B. B. Zaidan, A. Al-Haiqi, M. L. M. Kiah, M. Hussain, and M. Abdulnabi, "Evaluation and selection of open-source EMR software packages based on integrated AHP and TOPSIS," *J. Biomed. Inform.*, vol. 53, pp. 390–404, Feb. 2015.

[53] M. Migdadi, "Knowledge management enablers and outcomes in the small-and-medium sized enterprises," *Ind. Manage. Data Syst.*, vol. 109, no. 6, pp. 840–858, 2009.

[54] N. H. Hussain, T. Siti, M. Tengku, S. Fadzilah, and M. Noor, "Speech Input as an Alternative Mode to Perform Multi-touch Gestures," *Telkomnika*, vol. 16, no. 3, pp. 1367–1375, 2018.

[55] K. Olson, "An examination of questionnaire evaluation by expert reviewers," *Field Methods*, vol. 22, no. 4, pp. 295–318, 2010.

[56] F. I. Feng, "A study on school leaders' ethical orientations in Taiwan," *Ethics Behav.*, vol. 21, no. 4, pp. 317–331, 2011.

[57] A. S. Albahri, A. A. Zaidan, O. S. Albahri, B. B. Zaidan, and M. A. Alsalem, "Real-time fault-tolerant mHealth system: Comprehensive review of healthcare services, opens issues, challenges and methodological aspects," *J. Med. Syst.*, vol. 42, no. 8, p. 137, 2018.

[58] S. Guo and H. Zhao, "Fuzzy best-worst multi-criteria decision-making method and its applications," *Knowl.-Based Syst.*, vol. 121, pp. 23–31, Apr. 2017.

[59] H. A. AlSattar, A. A. Zaidan, B. B. Zaidan, M. R. A. Bakar, R. T. Mohammed, O. S. Albahri, M. A. Alsalem, and A. S. Albahri, "MOGSABAT: A metaheuristic hybrid algorithm for solving multi-objective optimisation problems," *Neural Comput. Appl.*, to be published.

[60] O. Enaizan, A. A. Zaidan, N. H. M. Alwi, B. B. Zaidan, M. A. Alsalem, O. S. Albahri, and A. S. Albahri, "Electronic medical record systems: Decision support examination framework for individual, security and privacy concerns using multi-perspective analysis," *Health Technol.*, to be published.

[61] M. A. O. Barrios, F. De Felice, K. P. Negrete, B. A. Romero, A. Y. Arenas, and A. Petrillo, "An AHP-topsis integrated model for selecting the most appropriate tomography equipment," *Int. J. Inf. Technol. Decis. Making*, vol. 15, no. 4, pp. 861–885, 2016.

[62] M. A. Ortíz, C. Nino, J. Patricia, A. De la Hoz, F. De Felice, and A. Petrillo, "An integrated approach of AHP-DEMATEL methods applied for the selection of allied hospitals in outpatient service," *Int. J. Med. Eng. Inform.*, vol. 8, no. 2, p. 87, 2016.

[63] N. M. Napi, A. A. Zaidan, B. B. Zaidan, O. S. Albahri, M. A. Alsalem, and A. S. Albahri, "Medical emergency triage and patient prioritisation in a telemedicine environment: A systematic review," *Health Technol.*, to be published.

[64] M. Iphar and S. Alpay, "A mobile application based on multi-criteria decision-making methods for underground mining method selection," *Int. J. Mining, Reclamation Environ.*, vol. 33, pp. 480–504, May 2018.

[65] J. Rezaei, "Best-worst multi-criteria decision-making method," *Omega*, vol. 53, pp. 49–57, Jun. 2015.

[66] O. S. Albahri, A. S. Albahri, A. A. Zaidan, B. B. Zaidan, M. A. Alsalem, A. H. Mohsin, K. I. Mohammed, A. H. Alamoodi, S. Nidhal, O. Enaizan, and M. A. Chyad, "Fault-tolerant mHealth framework in the context of IoT-based real-time wearable health data sensors," *IEEE Access*, vol. 7, pp. 50052–50080, 2019.

[67] E. M. Almahdi, A. A. Zaidan, B. B. Zaidan, M. A. Alsalem, O. S. Albahri, and A. S. Albahri, "Mobile patient monitoring systems from a benchmarking aspect: Challenges, open issues and recommended solutions," *J. Med. Syst.*, vol. 43, no. 7, p. 207, 2019.

[68] M. Alsalem, A. A. Zaidan, B. B. Zaidan, O. S. Albahri, A. H. Alamoodi, A. S. Albahri, A. H. Mohsin, and K. I. Mohammed, "Multiclass benchmarking framework for automated acute leukaemia detection and classification based on BWM and group-VIKOR," *J. Med. Syst.*, vol. 43, no. 7, p. 212, 2019.

[69] H. Gupta and M. K. Barua, "Supplier selection among SMEs on the basis of their green innovation ability using BWM and fuzzy TOPSIS," *J. Cleaner Prod.*, vol. 152, pp. 242–258, May 2017.

[70] H. Gupta and M. K. Barua, "A framework to overcome barriers to green innovation in SMEs using BWM and Fuzzy TOPSIS," *Sci. Total Environ.*, vol. 633, pp. 122–139, Aug. 2018.

[71] P. You, S. Guo, H. Zhao, and H. Zhao, "Operation performance evaluation of power grid enterprise using a hybrid BWM-TOPSIS method," *Sustainability*, vol. 9, no. 12, p. 2329, 2017.

[72] E. M. Almahdi, A. A. Zaidan, B. B. Zaidan, M. A. Alsalem, O. S. Albahri, and A. S. Albahri, "Mobile-based patient monitoring systems: A prioritisation framework using multi-criteria decision-making techniques," *J. Med. Syst.*, vol. 43, no. 7, p. 219, 2019.

[73] I. Masudin and M. G. F. Ayni, "Pengambilan keputusan multi kriteria: Kajian teoritis metode dan pendekatan dalam pemilihan pemasok," *J. Ilmiah Teknik Ind.*, vol. 17, no. 1, pp. 1–12, 2018.

[74] W. N. K. W. Ahmad, J. Rezaei, S. Sadaghiani, and L. A. Tavasszy, "Evaluation of the external forces affecting the sustainability of oil and gas supply chain using Best Worst Method," *J. Cleaner Prod.*, vol. 153, pp. 242–252, Jun. 2017.

[75] S. Ghaffari, A. Arab, J. Nafari, and M. Manteghi, "Investigation and evaluation of key success factors in technological innovation development based on BWM," *Decis. Sci. Lett.*, vol. 6, no. 3, pp. 295–306, 2017.

[76] O. Bataineh, D. A. Hjeelah, and S. Arabiat, "Multi-criteria decision making using AHP to select the best CAD software," in *Proc. Int. Conf. Intell. Inf. Technol. Ind.*, 2017, pp. 106–115.

[77] N. Jaini and S. Utyuzhnikov, "A fuzzy trade-off ranking method for multi-criteria decision-making," *Axioms*, vol. 7, no. 1, p. 1, 2017.

[78] M. A. Qader, B. B. Zaidan, A. A. Zaidan, S. K. Ali, M. A. Kamaluddin, and W. B. Radzi, "A methodology for football players selection problem based on multi-measurements criteria analysis," *Measurement*, vol. 111, pp. 38–50, Dec. 2017.

[79] N. Kalid, A. A. Zaidan, B. B. Zaidan, O. H. Salman, M. Hashim, O. S. Albahri, and A. S. Albahri, "Based on real time remote health monitoring systems: A new approach for prioritization 'large scales data' patients with chronic heart diseases using body sensors and communication technology," *J. Med. Syst.*, vol. 42, no. 4, p. 69, 2018.

[80] M. Alaa, I. S. M. A. Albakri, C. K. S. Singh, H. Hammed, A. A. Zaidan, B. B. Zaidan, O. S. Albahri, M. A. Alsalem, M. M. Salih, E. M. Almahdi, and M. J. Baqer, "Assessment and ranking framework for the English skills of pre-service teachers based on fuzzy Delphi and TOPSIS methods," *IEEE Access*, vol. 7, pp. 126201–126223, 2019.

[81] A. S. Albahri, O. S. Albahri, A. A. Zaidan, B. B. Zaidan, M. Hashim, M. A. Alsalem, A. H. Mohsin, K. I. Mohammed, A. H. Alamoodi, O. Enaizan, S. Nidhal, O. Zughoul, F. Momani, M. A. Chyad, K. H. Abdulkareem, K. A. Dawood, E. M. Almahdi, G. A. A. Shafeey, and M. J. Baqer, "Based multiple heterogeneous wearable sensors: A smart real-time health monitoring structured for hospitals distributor," *IEEE Access*, vol. 7, pp. 37269–37323, 2019.

[82] B. Margret, "Teaching speaking skills in English language to engineering students using language lab technology," in *Proc. Int. Conf. Technol. Educ.*, Jul. 2010, pp. 224–225.

[83] V. Ii, "Use of various activities to enhance speaking skill of engineering students," Tech. Rep., 2016.

[84] T. Read and A. Kukulska-hulme, "The role of a mobile app for listening comprehension training in distance learning to sustain student motivation," *J. UCS*, vol. 21, no. 10, pp. 1327–1338, 2015.

[85] A. P. Gilakjani and N. B. Sabouri, "Learners' listening comprehension difficulties in English language learning: A literature review," *English Lang. Teach.*, vol. 9, no. 6, pp. 123–133, 2016.

[86] S. Nazara, "Students' perception on EFL speaking skill development," *J. English Teach.*, vol. 1, no. 1, pp. 28–43, 2011.

[87] M. T. Yeganeh and P. Malekzadeh, "The effect of bilingualism on the developing of English reading skill," *Procedia-Social Behav. Sci.*, vol. 192, pp. 803–810, Jun. 2015.

[88] L. T. Tuan, "Enhancing EFL learners' writing skill via journal writing," *English Lang. Teach.*, vol. 3, no. 3, pp. 81–88, 2010.

[89] E. Faliyanti and M. Arlin, "The correlation between students' habit in watching English movie and vocabulary mastery at the sixth semester at English education study program in Muhammadiyah University of metro," *Intensive J.*, vol. 1, no. 2, pp. 94–102, 2018.

[90] A. Rodziah, S. M. Hasan, A. Rusli, and C. Noraini, "Integrated model for e-learning acceptance," in *Proc. IOP Conf. Ser., Mater. Sci. Eng.*, 2016, vol. 105, no. 1, p. 12016.

[91] N. Rhodin, "The influence of ICT on learners' motivation towards learning English," Malmö Univ., Malmö, Sweden, Tech. Rep., 2016.

[92] J. Lu, Z. Mao, M. Wang, and L. Hu, "Goodbye maps, hello apps? Exploring the influential determinants of travel app adoption," *Current Issues Tourism*, vol. 18, no. 11, pp. 1059–1079, Sep. 2015.

IEEE Access

N. K. Ibrahim *et al.*: Multi-Criteria Evaluation and Benchmarking for Young Learners' English Language Mobile Applications

[93] M. Talal, A. A. Zaidan, B. B. Zaidan, O. S. Albahri, M. A. Alsalem, A. S. Albahri, A. H. Alamoodi, M. L. M. Kiah, F. M. Jumaah, and M. Alaa, "Comprehensive review and analysis of anti-malware apps for smartphones," *Telecommun. Syst.*, to be published.

[94] D. Karabasevic, D. Stanujkic, S. Urosevic, G. Popovic, and M. Maksimovic, "An approach to criteria weights determination by integrating the Delphi and the adapted SWARA methods," *Manage. J. Sustain. Bus. Manage. Solution Emerg. Econ.*, vol. 22, no. 3, pp. 15–25, 2017.

[95] Y. Wu, "Applying of listening strategies in English as a foreign language," in *Proc. 5th Int. Conf. Law, Lang. Discourse*, 2015, p. 119.

[96] A. Khamkhien, "Teaching English speaking and English speaking tests in the thai context: A reflection from Thai perspective," *English Lang. Teach.*, vol. 3, no. 1, pp. 184–190, 2005.

[97] A. Raquela and J. E. Rini, "Cognitive domains found on speaking skill questions used in English language textbook," *Kata Kita*, vol. 4, no. 1, pp. 38–42, 2016. doi: 10.9744/KATAKITA.4.1.38-42.

[98] D. Wang, "Research on the new mode of college English teaching based on interactive teaching and flip-up classroom," in *Proc. 2nd Int. Conf. Econ., Manage. Eng. Educ. Technol. (ICEMEET)*, vol. 87, 2017, pp. 729–732.

[99] K. Hyland, "Faculty feedback: Perceptions and practices in L_2 disciplinary writing," *J. Second Lang. Writing*, vol. 22, no. 3, pp. 240–253, 2013.

[100] B. T. Wang, C. W. Teng, and H. T. Chen, "Using iPad to facilitate English vocabulary learning," *Int. J. Inf. Educ. Technol.*, vol. 5, no. 2, pp. 100–104, 2015.

[101] K. LeBeau, L. G. Huey, and M. Hart, "Assessing the quality of mobile apps used by occupational therapists: Evaluation using the user version of the mobile application rating scale," *JMIR mHealth uHealth*, vol. 7, no. 5, 2019, Art. no. e13019.

[102] R. Larsen, "Gamified mobile application to raise awareness and support energy efficient behavior," M.S. thesis, NTNU, Trondheim, Norway, 2018.

[103] D. D. M. Dolawattha, H. K. S. Premadasa, and P. M. Jayaweera, "Modelling the learner's perspectives on mobile learning in higher education," in *Proc. 18th Int. Conf. Adv. ICT Emerg. Regions*, 2018, pp. 133–141.

[104] O. S. Albahri, A. A. Zaidan, B. B. Zaidan, M. Hashim, A. S. Albahri, and M. A. Alsalem, "Real-time remote health-monitoring systems in a medical centre: A review of the provision of healthcare services-based body sensor information, open challenges and methodological aspects," *J. Med. Syst.*, vol. 42, p. 164, Sep. 2018.

[105] M. Daumer, U. Held, K. Ickstadt, M. Heinz, S. Schach, and G. Ebers, "Reducing the probability of false positive research findings by pre-publication validation–experience with a large multiple sclerosis database," *BMC Med. Res. Methodol.*, vol. 8, no. 1, p. 18, 2008.

[106] M. Dallachiesa, B. Nushi, K. Mirylenka, and T. Palpanas, "Uncertain time-series similarity: Return to the basics," *Proc. VLDB Endowment*, vol. 5, no. 11, pp. 1662–1673, 2012.

[107] E. Y. Mulyanah, I. Ishak, and R. K. Dewi, "The effect of communicative language teaching on students' speaking skill," *CICES*, vol. 4, no. 1, pp. 67–75, 2018.

[108] M. Bojovic, "Reading skills and reading comprehension in English for specific purposes," presented at the Annu. Meeting Eur. Bus. Document Assoc., Sep. 2010.

[109] I. N. Astawa, I. B. N. Mantra, and I. A. M. S. Widiastuti, "Developing communicative English language tests for tourism vocational high school students," *Int. J. Social Sci. Hum.*, vol. 1, no. 2, pp. 58–64, 2017.

[110] L. P. Ximenes, C. F. Guterres, and S. Pereira, "A study on the ability of 4th semester students' writing skill of english study program in the academic year 2018," *J. Innov. Stud. Character Educ.*, vol. 3, no. 1, pp. 155–164, 2019.

[111] S. Afrin, "Writing problems of non-English major undergraduate students in Bangladesh: An observation," *Open J. Soc. Sci.*, vol. 4, no. 3, p. 104, 2016.

[112] I. M. Alfaki, "Vocabulary input in English language teaching: Assessing the vocabulary load in spine five," *Int. J. English Lang. Linguist. Res.*, vol. 3, no. 1, pp. 1–14, 2015.

[113] Y. Benajiba, P. Rosso, L. Abouenour, O. Trigui, K. Bouzoubaa, and L. Belguith, "Question answering," in *Natural Language Processing of Semitic Languages*. Berlin, Germany: Springer, 2014, pp. 335–370.

N. K. IBRAHIM received the B.Sc. degree in computer science from the University of Basrah, Basrah, Iraq, in 2010, and the M.Sc. degree in artificial intelligence from Universiti Pendidikan Sultan Idris, Tanjung Malim, Malaysia. He works as a Teacher at a secondary school. His research areas are education application and artificial intelligence.

HAMSA HAMMED received the bachelor's degree in English language studies from the Faculty of Languages, Baghdad University, Baghdad, Iraq, in 2007, the master's degree in English language studies from the Faculty of Arts and Social Science, National University of Malaysia, and the Ph.D. degree in education (English language) from University Science Islamic Malaysia, in 2017. She is currently working as a Senior Lecturer with the Department of Early Childhood Education, Sultan Idris Education University. She led or has been a member of many funded research projects, and has published more than 20 articles at various international conferences and journals. Her research areas include English language studies, second language acquisition, learning English as a second language for young learners, and early childhood education.

A. A. ZAIDAN received the first-class B.Eng. degree in computer engineering from the University of Technology, Baghdad, Iraq, in 2004, the M.Sc. degree in data communications and computer network from the University of Malaya, Malaysia, in 2009, and the Ph.D. degree in artificial intelligence from Multimedia University, Malaysia, in 2013. He is currently working as a Senior Lecturer with the Department of Computing, University Pendidikan Sultan Idris. He led or has been a member for many funded research projects. He has published more than 150 articles at various international conferences and journals. His research areas are artificial intelligence, decision theory, data communication and network, AI applications on telemedicine, and E-health.

B. B. ZAIDAN received the B.Sc. degree in applied mathematics from Al-Nahrain University, Baghdad, Iraq, in 2004, and the M.Sc. degree in data communications and information security from the University of Malaya, Malaysia, in 2009. He is currently working as a Senior Lecturer with the Department of Computing, University Pendidikan Sultan Idris. He led or has been a member for many funded research projects, and has published more than 150 articles at various international conferences and journals. His research areas are artificial intelligence, decision theory, information security and network, and multicriteria evaluation and benchmarking.

O. S. ALBAHRI received the B.Sc. degree in computer science from Al Turath University College, Baghdad, Iraq, in 2011, the M.Sc. degree in computer science and communication from Arts, Sciences and Technology University in Lebanon, Beirut, Lebanon, in 2014, and the Ph.D. degree in artificial intelligence from Universiti Pendidikan Sultan Idris (UPSI), Tanjung Malim, Malaysia, in 2019. He led or has been a member for many funded research projects, and has published more than 20 articles at various ISI/WOS international journals. His research areas are decision theory, artificial intelligence, and medical informatics.

N. K. Ibrahim *et al.*: Multi-Criteria Evaluation and Benchmarking for Young Learners' English Language Mobile Applications

IEEE *Access*

M. A. ALSALEM received the B.Sc. degree in computer science from the University of Mosul, Iraq, in 2010, and the M.Sc. degree in Computer Science from the University of Mosul, in 2014. He is currently pursuing the Ph.D. degree with the Universiti Pendidikan Sultan Idris (UPSI), Tanjung Malim, Malaysia. He is currently working as a Lecturer with the University of Mosul. He led or has been a member for many funded research projects, and has published more than 25 articles at various international journals. His research areas are machine learning, telemedicine, and multicriteria decision making.

M. J. BAQER received the B.Sc. degree in software engineering from the Baghdad College of Economics Sciences University, Iraq, in 2013, and the M.Sc. degree in information technology from Universiti Pendidikan Sultan Idris (UPSI University), Malaysia, in 2018. His research area are multicriteria decision making.

R. T. MOHAMMED received the B.Sc. degree in computer science from the University of Baghdad, Iraq, in 2004, and the M.Sc. degree in Computer Science from the University of Technology-Baghdad, Iraq, in 2007. She is currently pursuing the Ph.D. degree with Universiti Putra Malaysia (UPM), Serdang, Malaysia. She is also working as a Senior Lecturer with the Geomatika University College, Malaysia. Her research areas are intelligent system, data mining, and multicriteria decision making.

SHAHAD NIDHAL received the B.Sc. degree in electrical and electronics engineering from the University of Technology, Iraq, in 1999, the M.Sc. degree in electrical engineering from UKM University, Malaysia, in 2005, and the Ph.D. degree from UKM Malaysia, in 2012. He is currently working as a Lecturer with MSU University. He led or has been a member for many funded research projects, and has published more than seven articles at various international conferences and journals. His research areas are pattern recognition, digital signal processing, signal processing, biomedical signal processing, and renewable energy.

ALI NAJM JASIM received the B.Sc. degree in software engineering from the University of Imam Jaafar Al-Sadiq, Baghdad, Iraq, in 2012, the M.Sc. degree in computer science information technology from Upsi University, Malaysia, in 2018, and graduated from the Universiti Pendidikan Sultan Idris (UPSI), Tanjung Malim, Malaysia. He is currently working as an Engineer with the Foundation of Alshuhda, Iraq. His research areas are education application, decision making, artificial intelligence, and medical informatics.

ALI H. SHAREEF received the B.Sc. degree in software engineering from the University of Imam Jaafar Al-Sadiq, Baghdad, Iraq, in 2013, the M.Sc. degree in computer science information technology from Upsi University, Malaysia, in 2018, and graduated from the Universiti Pendidikan Sultan Idris (UPSI), Tanjung Malim, Malaysia. He is currently working as a Lecturer with the University of Thiqar, Iraq. His research areas are education application, decision making, artificial intelligence, and medical informatics.

E. M. ALMAHDI received the B.Sc. degree in computer science from the University of Baghdad, Iraq, in 2000, and the M.Sc. degree in computer Science artificial intelligence from UPSI University, Malaysia, in 2018. He is currently a Student with the Universiti Pendidikan Sultan Idris (UPSI), Tanjung Malim, Malaysia. His research areas are medical informatics and decision making.

N. S. JALOOD received the B.Sc. degree in software engineering from the University of Imam Jaafar Al- Sadiq, Baghdad, Iraq, in 2013, the M.Sc. degree in computer science information technology from Upsi University, Malaysia, in 2018, and graduated from the Universiti Pendidikan Sultan Idris (UPSI), Tanjung Malim, Malaysia. He is currently working as a Senior Engineer with the Ministry of Education, Iraq. His research areas are education application, decision making, artificial intelligence, and medical informatics.

MUSAAB ALAA received the bachelor's degree in English in arts from the University of Al-Ma'mun, Baghdad, and the master's degree from the Department of Computer System and Technology, University of Malaya, Malaysia. He is currently pursuing the Ph.D. degree with the Department of Language and Communication, Universiti Pendidikan Sultan Idris, Malaysia, and teaching English as a second language.

● ● ●