Developing and Evaluating Gamifying Learning System by Using Flow-Based Model

Chung-Ho Su & Kai-Chong Hsaio
Shu-Te University, TAIWAN

Game-based learning is an effective learning method, whose performance depends on the quality of the educational game. Due to versatile game environments with complex backgrounds, evaluations are not easy to implement. Consequently, it is difficult for educators to determine to what degree a game may be qualified. This study proposes a novel, multiple-criteria decision making (MCDM) model based on Flow theory to help educators evaluate educational games. This study designs a Flow dimension questionnaire in which some domain experts are asked to re-examine the attributes of the Flow dimension by using the Delphi method. The extracted attributes of Flow dimension are then employed to build the proposed model for evaluating a given educational game. For practical implementation, this study developed an Evaluating Mobile-Learning Game System (EMGS), based on the proposed model. In the case study, several experts were requested to use the developed system for evaluating three different types of educational games. After the evaluation process, the experts were invited to fill out a questionnaire of system satisfaction. The results show that the developed system is faster and more convenient than the traditional evaluation method, and the developed system could effectively recommend the best qualified educational game for educators in different situations.

Keywords: game-based learning, Flow theory, multiple-criteria, decision making, Delphi method

INTRODUCTION

The increasing sophistication of information technologies has changed our lifestyles. Human beings – children to adults – live in a media-rich society with ubiquitous networks. Although, some children may be too young to start using digital technology products, digital games can be an important factor in attracting children to these products. In fact, ever-newer generations of digital games are indispensable for many people. Digital games seem to have a kind of "magic power," which makes people want to play them. This "magic power" has resulted in more and more educators attempting to develop game-based learning in education. In contemporary society, game-based learning has become a common means of education. The Horizon Report (Johnson, Smith, Willis, Levine & Haywood, 2011) indicated that game-based learning would be widespread in two to three years.

Hogle (1996) pointed out the benefits of educational games, such as stimulating motivation and interest, improving retention, effects of practice and feedback, and improving a higher order of skills.

Although, game-based learning has many teaching advantages, it poses the dilemma of effectively evaluating its own positive impact on learning. The learning effect depends on how much attention is paid to these game content interactions, guided by digital game-based learning (DGBL). However, the game environment is too complex to assess easily, and currently there are no unified standards to evaluate educational games. Thus, most educators evaluate the learning effect by using a pre-test before the game-based learning and a post-test at the end of the course. Evaluating a game-based learning system is not only time-consuming, but also suffers disturbance by many factors. Students may attempt to curry favor with the teacher by selecting positive items, even if the game is unattractive, which would have the converse effect of promoting a deficient game as successful and qualitative.

The Flow theory proposed by Csikszentmihalyi (1975) describes a state in which people are engaged in their favorite activities and nothing seems to matter. The Flow state occurs in many activities, such as rock-climbing, dancing, chess and playing games. In a Flow state, a player's potential ability is activated due to his/her intense focus on the task at hand. The goal of game-based learning is to lead students to this state, thereby, improving learning efficiency. Thus, if educators are unable to evaluate the Flow of complex game environments, than who can?

The relevant literature indicates that Flow has been widely studied, and that a number of Flow antecedents have been identified. However, most of the studies were based on a web environment, which is different from a game environment. Sweetser & Wyeth (2005) proposed a standard of game Flow for designing games with Flow, but it still lacked a unified evaluation criteria. Therefore, this study proposes a mobile-learning game model, based on Flow theory, for constructing a new Flow dimension with greater suitability for evaluating educational learning games.

The remainder of this paper is organized, as follows: section 2 reviews related works, including digital game-based learning, Flow theory, multiple-criteria decision making (MCDM), and the Delphi method; section 3 introduces an evaluation model, a system development and a proposed evaluation algorithm; section 4 is a case verification; and section 5 draws conclusions.

State of the literature

- The extracted attributes of flow dimension are employed to build the proposed model for evaluating educational game by using Delphi method, TOPSIS and OWA.
- Author develops three mobile serious games as an alternative evaluation solution for proposed flow-based MCDM evaluation model.
- The result show that developed system is more fast and convenient than traditional evaluation, and the developed system could effectively recommend the best qualified educational game for educators in different situations.

Contribution of this paper to the literature

- The author proposes a flow-based MCDM model for mobile game based learning system to evaluate gamifying learning system.
- Mobile-Learning Game System (EMGS) can help educators to aid teaching and enhance learning for student.
- Educators can easy to evaluate educational game by a novel MCDM model based on flow theory.

RELATED WORKS

This section briefly introduces digital game-based learning, Flow theory, MCDM, ordered weighted averaging, and the Delphi method.

Digital game-based learning

The definition of games, proposed by Salen & Zimmerman (2004), is "A system in which players engage in artificial conflict, defined by rules, that results in a quantifiable outcome." While digital games follow this definition they, further incorporate technology into the gaming systems. Studies have indicated that digital games can enhance many different positive effects, such as improving students' learning motivation (Kebritchi, Hirumi, & Bai, 2010), promoting in-depth learning and creative thinking (Eow, Ali, Mahmud & Baki, 2009), and providing powerful and meaningful contexts for learning (Shaffer, 2006). The effect promoted in education is called digital game-based learning.

In digital game-based learning, the learner engages in learning activities through solving problems or overcoming challenges proposed by games, while, simultaneously, being provided with a sense of achievement (Prensky, 2001). Specifically, learning accrues from actualizing the game's tasks, knowledge is enhanced through the game's content, and skills are improved while playing the game (McFarlane, Sparrowhawk & Heald, 2002).

The design of digital games is critical to DGBL. Successful digital games must incorporate a variety of characteristics, such as challenge, curiosity and fantasy to increase interest and intrinsic motivation for learning (Provenzo, 1991; Dickey, 2005). Also, continued game practice facilitates learner retention of information more easily (Dondi & Moretti 2007). Additionally, games should provide immediate feedback, enabling players to test their hypotheses and ascertain immediate results (Sung, Chang & Lee, 2008); and provide cognitive feedback by requiring learners to use previously learned skills or reminding learners of previously gained knowledge (Oblinger, 2004). The above reminder can be applied in Flow theory. The next part introduces Flow theory.

Flow theory

Csikszentmihalyi (1975) proposed the original definition of Flow as "the holistic experience that people feel when they act with total involvement." Flow describes a state of complete absorption or engagement in an activity, which is similar to an optimal experience (Csikszentmihalyi & Kleiber, 1991). During this state, people are so involved in a goal-oriented activity that nothing else seems to matter. Because the experience is so pleasant, people want to engage and reengage in the same activities. The design aim of learning games is to replicate this experience for the learner, facilitating unconscious learning by heart.

The literature related to the concept of Flow has been extensively studied. Finneran & Zhang (2003) proposed a person-artifact-task (PAT) model, which conceptualized the major components of people working on computer-mediated activity. According to the model, the occurrence of Flow depends on the interaction among the people, task and artifact. This model reminds us of what really influences the Flow experience. Flow roughly is divided into three phases: Flow antecedents, Flow state (Flow experience) and Flow consequences as Figure 1.

Flow antecedents

One of the purposes of this study is to select the appropriate Flow dimension (antecedents) for evaluating educational mobile-learning games. The possibility of a game's Flow depends on the quantity and quality of the flow antecedents. Many Flow antecedents have been confirmed in the related literature, such as: clear goals and appropriate feedback (Chen, Wigand & Nilan, 1999); telepresence (Hoffman, & Novak, 1996; Novak, Hoffman & Yung, 2000); a perception of challenges, which matches the player's skills or knowledge (Chen et al. 1999); ease of use (Hsu & Lu, 2003); novelty (Huang, 2003); personal innovativeness (Agarwal & Karahanna,
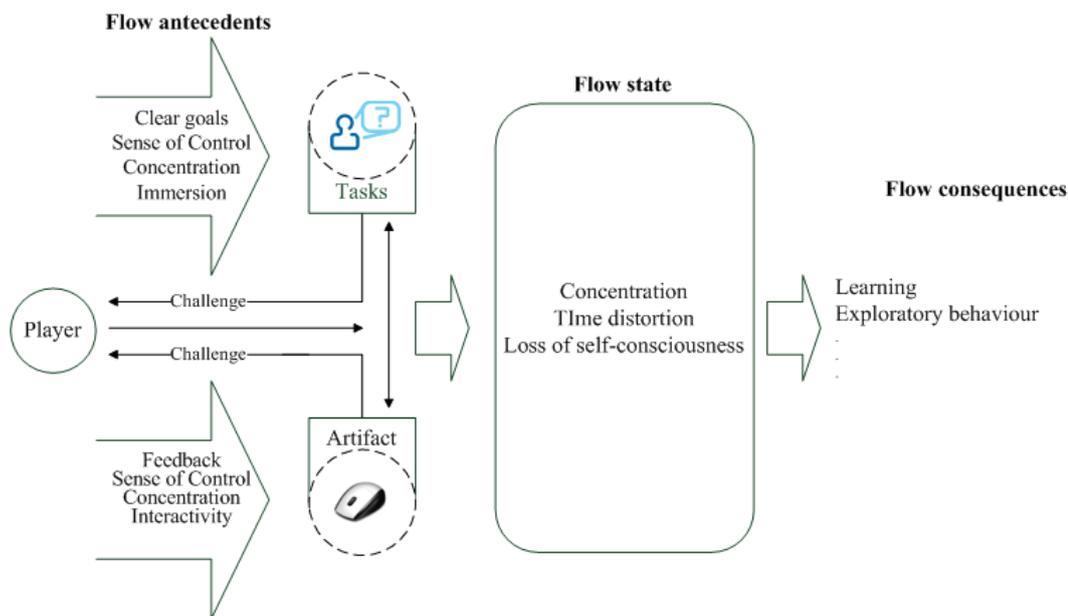
Figure 1. The flow framework

2000); and attractiveness (Skadberg & Kimmel, 2004). Although the above Flow antecedents have been confirmed, a large number of dimensions are not conducive to multiple-criteria decision making. Thus, this study only selects the antecedents adopted by most studies. A detailed description of the selected dimensions is presented in Section 3.

Flow state

The Flow state is comprised of concentration, loss of self-consciousness and time distortion. During the Flow state, a person is so totally focused on playing the game that s/he forgets everything, including unpleasant things, while the passage of time goes unnoticed. Because Flow activities require a complete concentration of attention to the task at hand, no remaining cognitive resources can be left to irrelevancies. Thus, self-awareness seems to disappear during Flow, and the actual passage of time does not seem to occur.

Flow consequences

Learners can learn more efficiently when there is complete concentration of attention on the learning objectives. Thus, the game must ensure that the learning objectives are related to the task. In addition, a game with Flow will promote intense curiosity, enabling learners to explore more possibilities in their quest for knowledge.

Multiple-Criteria decision making

In the real world, human beings face many decision making problems, including ones with conflicting criteria. Multiple-criteria decision making (MCDM) was developed to provide assistance to decision makers in choosing among alternatives. According to the literature, MCDM can be divided into two types: Multi-Attribute Decision Making (or MADM) and Multi-Objective Decision Making (or MODM). Generally speaking, Multiple Objective Decision-Making (MODM), as a continuous evaluation, is a model planned with mathematics in order to acquire the alternative of a decision. Multiple Attribute Decision-Making (MADM), on the other hand, is a discrete evaluation mainly defining the optimal alternative among various ones by evaluating the relative importance of attributes. In the Multiple Attribute Decision-Making model, a decision-maker evaluates several alternatives, under various

objectives or attributes, to determine the priority (Hwang and Yoon, 1981); while MADM studies decision problems in which the decision space is continuous. In these decision making situations, the decision alternatives have been predetermined. Most studies are based on MCDM, for which an optimal solution is relatively easy to obtain.

MCDM has been applied to many fields, including mathematics, behavioral decision theory, economics, computer technology, software engineering and information systems (Behzadian, Otaghsara, Yazdani & Ignatius, 2012). A considerable number of decision making models has been developed based on the MCDM theory, such as the weighted sum model (WSM), with the earliest and probably the most widely used being the analytic hierarchy process (AHP, Saaty, 1980), and a technique for Order Performance by Similarity to Ideal Solution (TOPSIS, Hwang & Yoon, 1981). Some recent methods can be listed as follows: Weighted Sum Method (Blanc & Jelassi,1989; Morisio,Tsoukis & IusWare,1997), TOPSIS (Mao, Mei & Ma,2009). In this study Delphi and TOPSIS methods are integrated and extended so as to utilize OWA weight ranking for the evaluation of the alternatives. Although the literature does not provide studies that directly focus on Game-based learning system selection; however MCDM methods are widely used in the field of information systems selection and this is also the niche of this study. This paper selects TOPSIS as the decision making method, the detailed content of which is introduced below.

TOPSIS

TOPSIS(Technique for Order Preference by Similarity to Ideal Solution), a method of Multiple-Criteria Decision Making, was proposed by Hwang & Yoon (1981).TOPSIS aims to search an alternative closest to Positive Ideal Solution and farthest Negative Ideal Solution. Positive Ideal Solution refers to the attribute with the maximum benefit or the minimum cost in the alternatives. On the contrary, an attribute with the minimum benefit and the maximum cost is regarded as Negative Ideal Solution, i.e. to pick the best apple among a barrel of rotten apples (Kua-Hsin Peng & Gwo-Hshiung Tzeng, 2013). Such a characteristic could help the decision be close to the ideal solution. The advantages of TOPSIS are: (1) TOPSIS logic is easily understood; and (2) the calculation processes are straightforward, and they also avoid the situation in which both the shortest distance from the positive ideal solution and the shortest distance from the negative ideal solution are unable to make decisions. TOPSIS allows a decision-maker giving weights according to the subjective value and preference. A decision-maker ranks attributes based on the importance and confirms the preference of such attributes to further decide the relative weights. Although weights are subjectively given by a decision-maker, they could be determined through group discussions.

In practice, TOPSIS has been successfully applied to solve selection/evaluation problems with a finite number of alternatives (Jee & Kang, 2000; Yong, 2006) because it is intuitive and easy to understand and implement. Furthermore, TOPSIS has a sound logic that represents the rationale of human choice (Shih, Syur & Lee, 2007) and has been proved to be one of the best methods in addressing the issue of rank reversal (Zanakis, Solomon, Wishart & Dublish, 1998). In this case, TOPSIS method presents better evaluation abilities to subjectively, objectively, and flexibly select adaptive decisions.

Rank reversal is a phenomenon in which the associated addition of a new alternative transforms an earlier, originally non-optimal alternative into an optimal one. At the same time, the theoretical development of the proposed alternative method avoids rank reversal. García-Cascales & Lamata (2012) believe that the reason for rank reversal is due to the selection of the best alternative, which depends on the other alternatives, called "relative mode." Thus, in order to avoid

this phenomenon, the alternative must be non-dependent, referred to as an "absolute mode." The TOPSIS procedure consists of the following steps:

Step 1: Establish a performance matrix: In the performance matrix, m is the number of alternatives, n is the number of criteria, and x_{mn} represents the performance values of the alternatives.

$$D = \begin{Bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{Bmatrix} \tag{1}$$

Step2: Normalize the decision matrix: The normalized performance matrix can be obtained by using the following formula:

$$N_{ij} = \frac{x_{ij}}{Max_i(x_{ij})} \ , \quad i = 1, \cdots, m \quad j = 1, \cdots, n \tag{2}$$

Step3: Calculate the weighted, normalized decision matrix: The weights can be obtained in different ways, such as ordered weighted averaging (OWA), AHP, etc. The weighted, normalized value is calculated as:

$$v_{ij} = w_j \cdot N_{ij}, \qquad i = 1, \cdots, m \quad j = 1, \cdots, n \tag{3}$$

Step 4: Determine the positive and negative ideal solutions: The positive ideal value, set at A^+, and the negative ideal value, set at A^-, are determined, and, moreover, to avoid the rank reversal the best alternative, v_b^+, and worst alternative, v_w^+, are added, as follows:

$$A^+ = \{v_1^+, \cdots, v_n^+, v_b^+\} = \{(max_i v_{ij} | j \in J)(min_i v_{ij} | j \in J')\} \ \ i = 1,2, \cdots, m \tag{4}$$

$$A^- = \{v_1^-, \cdots, v_n^-, v_w^+\} = \{(min_i v_{ij} | j \in J)(max_i v_{ij} | j \in J')\} \ \ i = 1,2, \cdots, m \tag{5}$$

Step 5: Calculate the separation measures:

This step distinguishes each alternative from the positive ideal solution (PIS), A^+, and the negative ideal solution (NIS), A^-, with an m-multidimensional, Euclidean distance employed to calculate the separation measures, defined, as follows:

$$D_i^+ = \left\{ \sum_{j=1}^n (v_{ij} - v_j^+)^2 \right\}^{\frac{1}{2}}, \ \ i = 1, \cdots, m \tag{6}$$

$$D_i^- = \left\{ \sum_{j=1}^n (v_{ij} - v_j^-)^2 \right\}^{\frac{1}{2}}, \ \ i = 1, \cdots, m \tag{7}$$

Step 6: Calculate the relative closeness to the ideal solution: The integrated value (I_i) to the ideal solution can be expressed, as follows:

$$I_i = \frac{D_i^-}{D_i^+ + D_i^-}, \quad i = 1, \cdots, m \tag{8}$$

$$if \ \bar{R}_i = 1 \rightarrow A_i = A^+$$

$$if \ \bar{R}_i = 0 \rightarrow A_i = A^-$$

The $\bar{I}_i$ value is between 0 and 1. The closer the I_i value is to 1, the higher the priority of the alternative.

Step 7: Rank the preference ordering. Rank the alternatives according to I_i with their values in descending order.

OWA

Yager (1988) proposes an ordered, weighted averaging (OWA), which has the ability to get the optimal weights of the attributes, based on the rank of these vectors after processing the aggregation. It is useful for multiple-criteria decision making, because MCDM problems often require the inclusion of information about the importance associated with the different criteria (Yager, 2004).

An OWA of dimension n is a mapping $f: R^n \rightarrow R$ that has an associated weighted vector $W = [w_1, w_2, \cdots, w_{1n}]^T$ with the following properties:

$$W_i \in [0,1] \quad \text{for} \quad i \in I = \{1,2,\cdots,n\} \text{ and } \sum_i w_i = 1$$

such that

$$f(a_1, a_2, \cdots, a_n) = \sum_{i=1}^{n} w_i b_i \tag{9}$$

where b_i is the ith largest element in the collection of aggregate objects $\{a_1, a_2, \cdots, a_n\}$.

Fullér & Majlender (2001) proposed a new OWA step based on a maximum entropy; it is a simplified step, based on the original OWA, in which the weight can be calculated by using only situational variable α, number of attributes n, and the importance ordering factor. Their step is defined as

$$Orness(W) = \frac{1}{n-1}\sum_{i=1}^{n}(n-i)\,W_i \tag{10}$$

The entropy of information is defined as

$$Disp(W) = -\sum_{i=1}^{n} W_i \ln W_i \tag{11}$$

O'Hagan, Palin & Davis (1988) combine the principle of maximum entropy and OWA to propose a particular OWA which has maximum entropy with a given level of orness. The definition is presented, as follows:

Maximize:

$$-\sum_{i=1}^{n} W_i \ln W_i$$

Subject to:

$$\alpha = \frac{1}{n-1}\sum_{i=1}^{n}(n-i)\,W_i \qquad 0 \leq \alpha \leq 1 \tag{12}$$

Where w_i is the weight vector, n is the number of attributes, and α is the situation parameter. According to their approach, when obtaining the value of n and α, the optimal value of w_1 can obtained by (14). Once w_1 is computed, then w_n can be determined from (15) and other weights obtained from (13).

$$\ln W_j = \frac{j-1}{n-1}\ln w_n + \frac{n-j}{n-1}\ln w_1 \Rightarrow W_j =$$

$$\sqrt[n-1]{W_1^{n-j} W_n^{j-1}} \tag{13}$$

and

$$W_1[(n-1)\alpha + 1 - nW_1]^n = \left((n-1)\alpha\right)^{n-1}[((n-1)\alpha - n)W_1 + 1] \tag{14}$$

If $\quad W_1 = W_2 = \cdots = W_n = \frac{1}{n} \Rightarrow disp(W) = \ln n$

then

$$Wn \quad \frac{((n-1)\alpha - n)W_1 + 1}{(n-1)\alpha + 1 - nW_1} \tag{15}$$

Delphi method

The Delphi method is one of the most commonly used qualitative forecasting techniques (Anderson, Sweeney & Williams, 1998). It is used to solve real world decision making problems systematically. RAND Corporation originally developed it during the 1950s and 60s as a primary way of allowing a group of experts to generate discussion and make policy decisions (Goodman, 1987). Afterwards, The Delphi method was defined as "A method for the systematic collection and aggregation of informed judgments from a group of experts on specific questions or issues" (Reid, 1988). This definition indicates that the Delphi method is suitable for group decision making, for gathering expert opinions, and for reiterating, with feedback, until reaching a consensus or a pre-defined criteria about a specific question or issue (Goodman, 1987). From the literature (Li, 2005; Turban & Aronson, 2001), it is obvious that decision making in groups has the following advantages: groups are better than individuals at catching errors and understanding problems; and groups have more information or knowledge. The Delphi method has thus been extensively used in various fields, such as education (Clayton, 1997), linguistic criteria evaluation (Cheng & Lin, 2002), library management (Hsieh, Chin & Wu, 2006), and other fields (Adler, 1996; Fish & Busby, 1996).

Based on previous studies (Linstone & Turoff, 1975; Goodman, 1987; Murry & Hammons, 1995; Webler, Levine, Rakel & Renn, 1991), the Delphi process can be described by the following step:

Step.1 clearly defines the problem and develops a questionnaire with a structured process concerning the problem;

Step.2 organizes a group of participants (about 8-15 experts), extending anonymity to each;

Step.3 asks participants to complete a questionnaire, and gathers the results;

Step.4 these results are then returned to each participant. If his/her views differ from the consensus, s/he will be asked to reevaluate his/her opinions and modify them, if wanted; and

Step.5 repeats these processes until a predefined stop criterion (e.g., stability of results, achievement of consensus, or number of rounds) emerges.

PROPOSED MODEL

This section describes the research concept and the building of an EMGS evaluation model, which includes a selected Flow variable, a screened variable by the Delphi method, and a proposed algorithm.

Research concept

In the past, it was difficult to determine whether a learning game was a qualified learning system. Most of the studies using a pre-test and a post-test to understand the course's learning effects were time-consuming and inaccurate. Although the Flow theory evaluation process has improved, there are still two problems: 1) Flow does not have a unified standard; and 2) there are no methods to assist educators in selecting educational games.

To overcome the above problems, this study proposes an evaluating, mobile-learning game model, based on an MCDM with Flow theory, to evaluate qualified

DGBL games. Several Flow dimensions items (variables) will be selected from past studies, and screened by experts based on the Delphi method. After determining the Flow dimensions, a game evaluation system EMGS will be developed to provide a complete evaluation process for educators. The process of the proposed model is shown in Figure 2.

Building DGBL evaluation model

The main purpose for constructing a DGBL evaluation model is to identify Flow dimensions. In a variety of contexts, one way or another, most people have experienced Flow. However, translating this intuitive understanding into a consistent definition has proven to be challenging (Hoffman & Novak, 2009). Even now, the developed DGBL evaluation models still lack a consistent concept of Flow. Thus, several Flow dimensions with evaluation items will be selected form past studies.

Selected Flow variable

In order to make the Flow dimension more suitable to evaluate educational games, this study examined the numerous antecedents of Flow from past studies and selected eight antecedents with 41 items which have been used in at least three studies. The antecedents include: a clear goal (5 items), feedback (5 items), challenge and skills (5 items), sense of control (5 items), concentration (6 items), immersion (5 items), interactivity (7 items), and knowledge improvement (3 items). The Flow dimension is shown in Table 1.
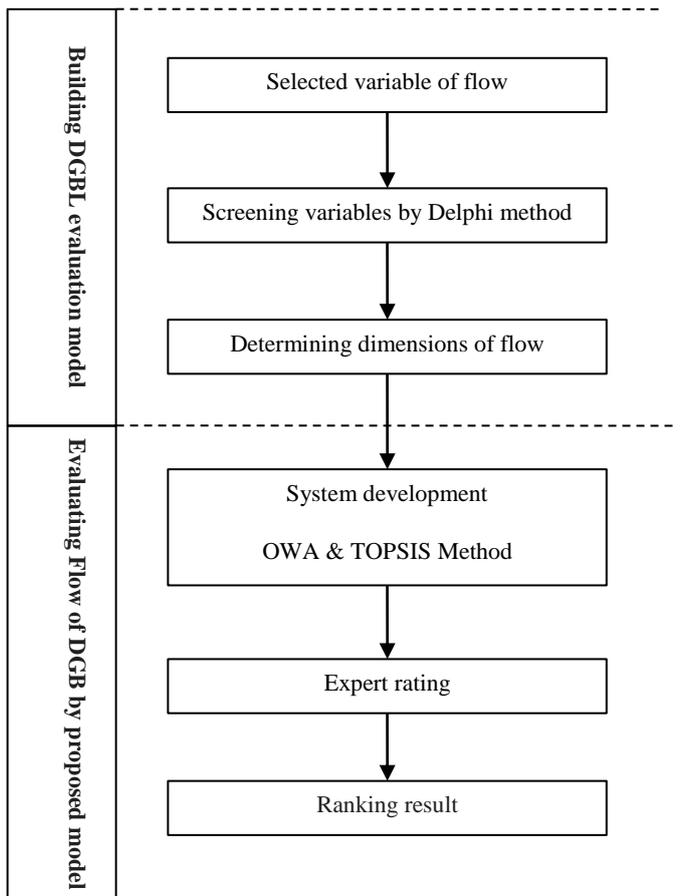
Figure 2. Process of the proposed model

Table 1. Flow dimensions with items and related literature

Dimensional	Item No.	Related literature
Clear goals	CG1, CG2, CG3, CG4, CG5	Csikszentmihalyi (1975);Federoff (2002);Sweetser & Wyeth (2005)
Feedback	F1, F2, F3, F4, F5	Csikszentmihalyi (1975);Federoff (2002);Sweetser & Wyeth (2005)
Challenge and skills	CS1,CS2,CS3,CS4,CS5	Csikszentmihalyi (1975);Hoffman & Novak (2000);Federoff (2002)
Sense of control	SC1,SC2,SC3,SC4,SC5	Hsu & Lu (2003);Federoff (2002);Sweetser & Wyeth (2005)
Concentration	C1,C2,C3,C4,C5,C6	Skadberg & Kimmel (2004);Federoff (2002);Sweetser & Wyeth (2005)
Immersion	I1,I2,I3,I4,I5	Hoffman & Novak, (2000);Novak, Hoffman & Yung (2000);Sweetser & Wyeth (2005)
Interactivity	IT1,IT2,IT3,IT4,IT5,IT6,IT7	Hoffman & Novak (2000);Novak, Hoffman & Yung (2000);Skadberg & Kimmel (2004);Huang (2003);Choi, Kim & Kim (2007);Sweetser & Wyeth (2005)
Knowledge improvement	K1,K2,K3	Fu, Su & Yu (2009)

Clear goals: The goals of the game are clear, and more easily keep the player's attention focused on the learning tasks. Although some games allow for curious exploration, it is still good practice to provide a clear, main goal in the beginning of an educational game. There should be several sub-goals, which guide the player towards the completion of tasks in order to achieve the game's end-goal. If the sub- or end-goal is too challenging, the probability of experiencing Flow is low. Furthermore, the goal of an educational game should be related to the course's learning objectives. If the learning objectives are separate and distinct from the game, then learning will be ineffective.

Feedback: The purpose of feedback is to inform the player of his/her performance, which allows players to control their own progress. Feedback can be subdivided into immediate feedback and cognitive immersion (Butler & Winne 1995). The game should provide immediate feedback to players. Immediate feedback sustains attention and avoids player distraction and loss of task focus, which could be caused by delayed feedback. Moreover, adding the appropriate prompt to feedback can prevent anxiety. Cognitive feedback is related to cognitive problem-solving, and repetition reminds players of the learning objectives, while aiding them in retaining skills and knowledge.

Challenges and skills: Challenges and skills are important dimensions. Generally, the aim of a learning game is to provide the learner with knowledge that can be learned through problem solving. Thus, the challenges must be in balance with the skill level of the player, and should be related to the main game task so that a Flow state is possible. The challenges and skills can be explained by the three-channel Flow model. The letter S represents four situations in which people engage in activities. At the beginning (S1), the player only has basic skills and the game provides the initial challenge. A player's skills increase when s/he engages in activities with Flow, which occurs when s/he feels that the game's difficulty is compatible with his/her skill set. Once the player's skills improve, s/he will feel bored and out of Flow (S2). Conversely, if skills do not improve, s/he might feel anxious because the game's challenges are too high (S4). Neither anxiety nor boredom is conducive to Flow. Thus, the challenges should vary, in accordance with a player's skills, and match the skill level of each individual player. In addition, game-supplied tips can improve the player's skills slightly, and enhance the possibility of the occurrence of Flow (zone of proximal development).

Sense of control: It is important that the learner feels a sense of control, which means the learner's cognitive resources are sufficient to bear the burden of the game's cognitive load. When the use of an artifact is complex, the artifact will divert

the learner's attention from the task. Poorly constructed interface decreases the likelihood of experiencing task-based Flow because the player has to sacrifice attention and other cognitive resources to inappropriate activities. Thus, the game interface must be simple and easy to use, one which can maintain Flow and keep the learner focused on the task. Further, the game play should not be too complex, but allow the learner to use strategies freely and play the game in a way that maximizes concentration.

Concentration: The aim of a learning game is to attract the attention of a learner while enhancing the occurrence of Flow. Thus, games should have appealing elements to keep learners focused, and they must ensure that learners are not distracted from their tasks. Furthermore, the main task of a learning game is that it must be highly relevant to the learning objectives.

Immersion: Immersion is a phenomenon similar to, and often confused with, Flow experience. Kiili, de Freitas, Arnab & Lainema (2012) explained the difference between these two. Immersion can be defined as a sensation of being surrounded by a completely other reality taking over all of our attention, but the completely concentrated attention of Flow tends to be more goal-oriented. Immersion is like a lower-level expression of a Flow state, and includes several important factors which should be considered during game design.

Interactivity: Social leaning theories raise the point that we learn through interactions with others. Educational games should provide a channel of communication between learners and others. Some educational games with role-playing allow learners to learn through conversation with non-player characters.

Knowledge improvement: It is undeniable that the main purpose of education games is to enhance learning. Tiger (2000) reminded us that an improvement in knowledge can be considered an enjoyable experience, conferred by an e-learning game onto a player. Thus, knowledge improvement is added to the Flow dimension as a special item.

Screening variables by the Delphi method

The many selected antecedents of Flow have been confirmed in past studies. However, too many variables increase the complexity of the evaluation process and also of multiple-criteria decision making. Therefore, the selected antecedents were screened by experts, using the Delphi method.

Experts selection

Nine experts were invited, including: three educators with experience in digital game-based learning; and six game experts with more than ten years of gaming experience. The rationale behind the number of game experts being greater than the number of educators is that the game was used more to raise interest than to improve knowledge. More game experts could build the dimensions, which focused on the attractiveness of the game, rather than the improvement of knowledge of the game. Further, each expert was well aware of Flow theory before the discussion of the Delphi method, and examined these variables to determine whether they could appropriately evaluate Flow from a different perspective.

The Delphi method

The questionnaire, designed by Likert and based on a five-point scale, proceeded, as follows: experts examined the appropriateness of an item, based on evaluating the Flow and the completed the questionnaire; after the first round, the data was collected, and the items with large standard deviations were regarded as inappropriate; inappropriate items were repeatedly examined by the experts until the experts were in agreement.

Furthermore, after the Delphi process, in order to obtain the degree of importance necessary for a Flow dimension, the experts were asked to rank the Flow dimension based on educational games. This important ordering will be used to calculate the OWA weight in this study.

Determine Flow dimensions

Through succeeding rounds, the Delphi process reaches results consistent with the experts' threshold. Eight dimensions with 25 items were retained. 17 inappropriate items were deleted, based on the consistency of the experts' threshold (standard deviation is 0.87), the threshold value retained the important items. The results and importance ordering (from 1 to 8, most to least important) are found on Table 2.

Proposed method

In order to identify the best mobile learning games, this study developed an evaluation system (an EMGS), based on the proposed model. In the process, this system integrated game description, play game, evaluation and ranking games. The system development and rank method are described in the following paragraphs.

TOPSIS was selected, based upon the concept that a chosen alternative should be the shortest distance from the Positive Ideal Solution (PIS) and the farthest from the Negative Ideal Solution (NIS). The advantage of TOPSIS is easily understood, a simple calculation and can assist decision makers in making decisions quickly.

Additionally, the weight of a Flow dimension is obtained by an OWA method. To obtain the weights, the number of attributes obtained by a number of Flow dimension and situation parameter α will be adjusted based on the results. The process of an OWA & TOPSIS method is shown as Figure 3

This study uses an OWA & TOPIS method, consisting of seven steps, presented as follows:

Table 2. Dimensions of flow after screening

Importance ordering	Dimension No.	Dimension	Item No.	Mean	Standard deviation
4	D1	Clear goals	CG3	4.444	0.726
			CG5	4.222	0.667
5	D2	Feedback	F1	4.667	0.500
			F2	4.556	0.527
			F4	4.444	0.527
			F5	4.111	0.782
3	D3	Challenge and skills	CS2	4.556	0.527
			CS3	4.667	0.500
			CS4	4.111	0.601
			CS5	4.556	0.527
1	D4	Sense of control	SC1	4.667	0.500
			SC2	4.556	0.527
			SC3	4.556	0.726
			SC4	4.667	0.500
			SC5	4.778	0.441
2	D5	Concentration	C1	4.667	0.500
			C2	4.667	0.500
			C3	4.667	0.500
			C4	4.000	0.707
6	D6	Immersion	I1	4.556	0.527
			I4	4.556	0.527
8	D7	Interactivity	IT3	4.222	0.833
			IT7	3.556	0.726
7	D8	Knowledge improvement	K1	3.778	0.833
			K3	4.111	0.782
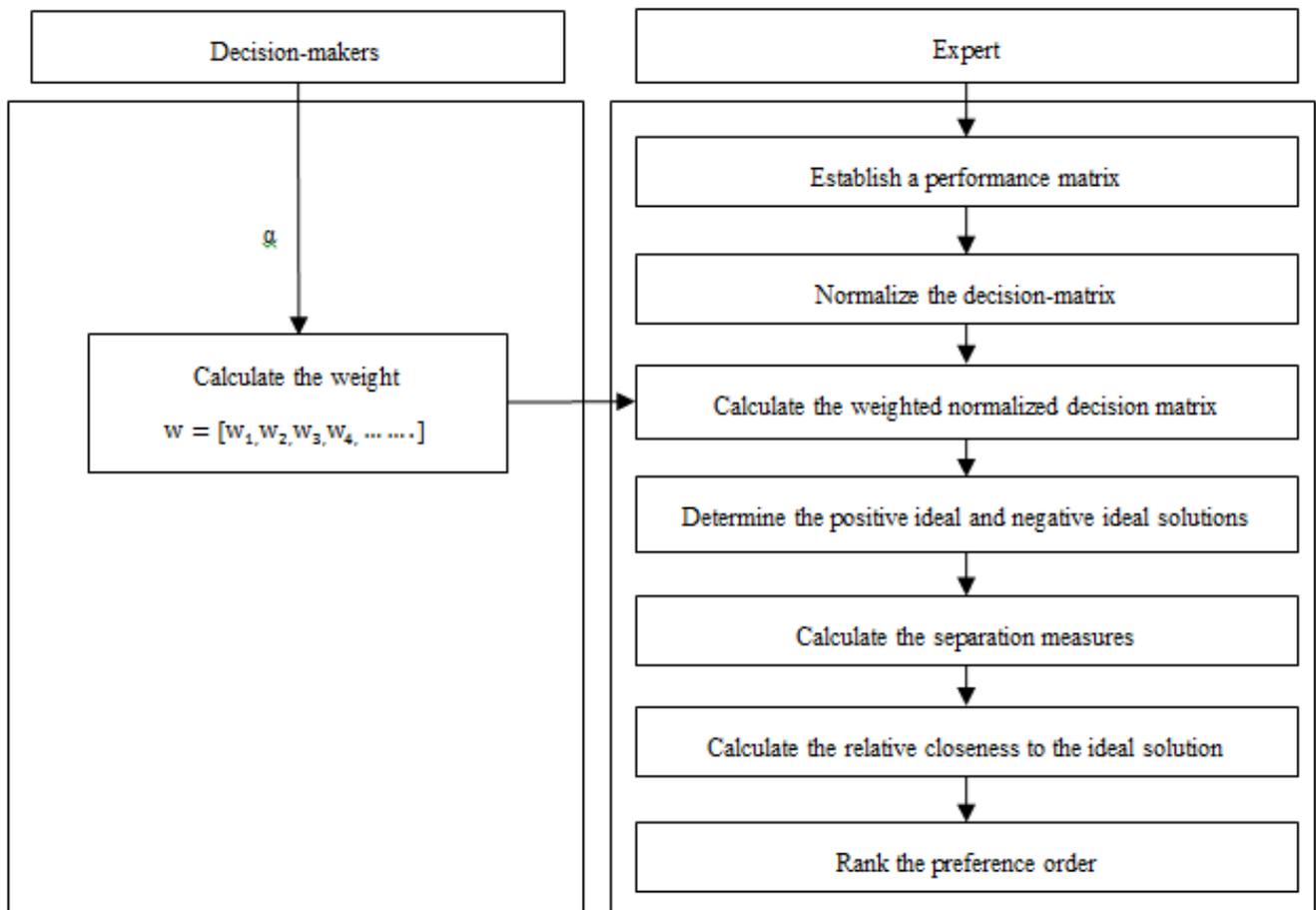
Figure 3. The process of OWA & TOPSIS method

Step 1: Establish a performance matrix
The performance matrix is obtained by experts evaluating the developed system, the scores of which include n Flow dimensions in m learning game.

Step 2: Normalize the decision matrix
In this study, we use equation (2) to normalize the decision matrix.

Step 3: Calculate the weighted normalized decision
In order to obtain the weight of the Flow dimension by an OWA method, the N parameters, the α within the N, and the important ordering of the Flow dimension must first be set. N is the number of Flow dimensions, α is a situational value, which can be decided by decision makers, and the importance ordering of the Flow dimension is, as per the experts' opinions. In this study, parameter N is 8, α is 0.6 (default), and the importance ordering, based on the experts' opinions, is (dimension 4> dimension 5> dimension 1> dimension 3> dimension 2> dimension 6> dimension 8> dimension 7).

Step 4: Determine positive ideal and negative ideal solutions
Positive and negative ideals are defined by equations (4) and (5).

Step 5: Calculate the separation measures
This step uses the m-multidimensional Euclidean distance of equation (6) and (7) to calculate distance.

Step 6: Calculate the relative closeness to the ideal
From equation (8), the integrated value (I value) can be gotten from the above steps. The integrated value "I" is between zero and one, and the closer to 1 the higher the priority of the alternative.

Step 7: Rank the preference order

The I value will be used to rank educational, mobile-learning games. The ranking results are shown in the system.

SYSTEM DEVELOPMENT

This section develops an evaluating mobile-learning game system, EMGS, based on the proposed model. The developed system combines game description, play game, rating game, and views the ranking result in an evaluation process.

The developed system has two main functions: rating games and ranking results. Rating a game includes viewing the introduction of the game, playing the game and rating the game. In order to compare the differences among games, the three or more alternatives (games) are organized as individual cases. Raters read the introduction to the games and play the games after they select an individual case. When a rater has a full understanding of the game, s/he can start to rate the game (Figure 4). The rater evaluates the game based on our evaluation model. The model includes eight Flow dimensions with 2 to 8 items, totaling 25 items. The dimension scores are averaged from each item score it contains. The item scores are between 1 and 10 points, the higher, the better. It is most important that, in order to compare the differences, raters rate each item of each game at the same time.

After the rating process, the rater can view the results immediately at rank result. The developed system shows the game's rank result and dimension scores (Figure 5). The game's rank is sorted by the integrated value - I value, which is calculated by
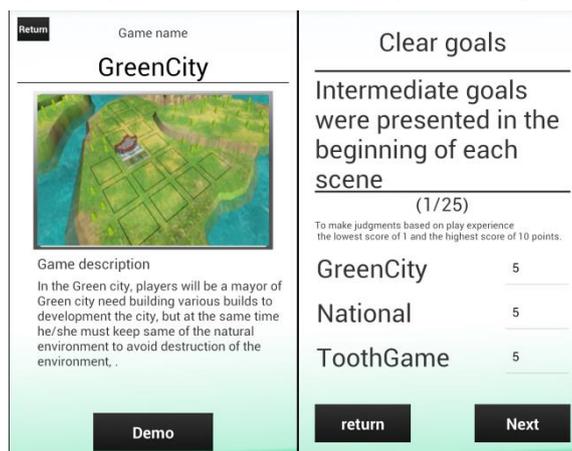
Figure 4. System interface description (1)
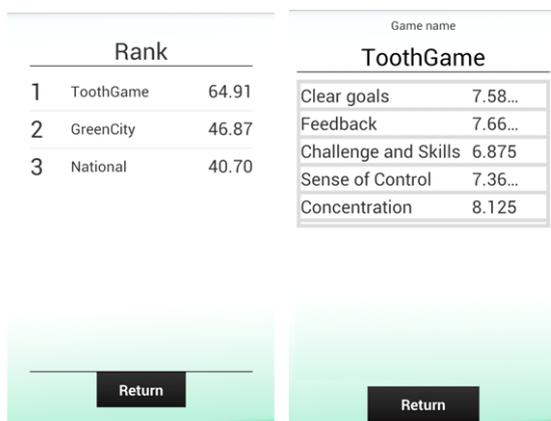
Figure 5. System interface description (2)
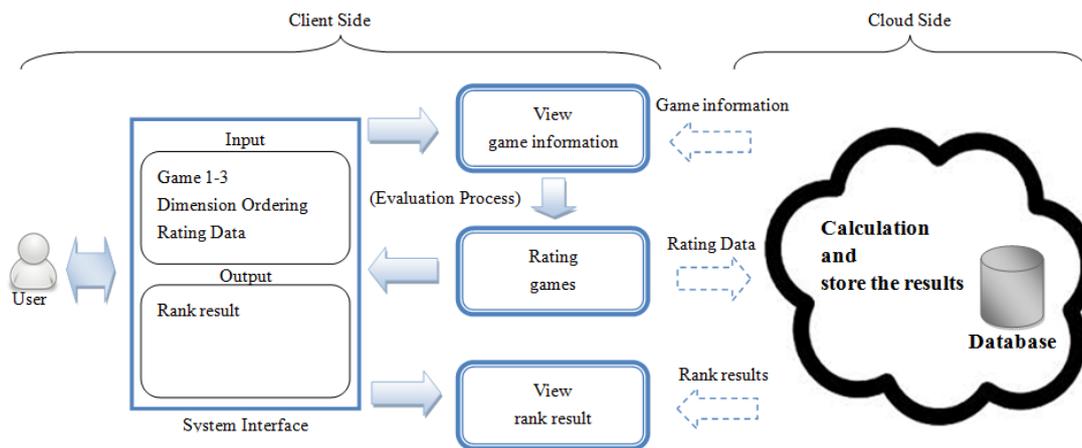
Figure 6. System architecture

an OWA & TOPSIS method with dimension scores. In addition, raters can adjust the importance ordering (weight of dimensions) according to the situation.

The developed system is an application software developed by Java for Smartphone. The system does not store any data locally, but calculates and stores data as with cloud services. The system architecture is shown in Figure 6.

CASE VERIFICATION

In this section, we use a case study to verify the proposed evaluation model. Three education games are evaluated by several experts using the developed EMGS system. The game description, participants and evaluation process, evaluation result and system satisfaction will be introduced in the following sections.

Game description

Three games with different types of knowledge and style were chosen from which to obtain data from experts. Game 1 is a green energy educational game, named GreenCity, Game 2 is a game involving a description of cardiac catheterization, named National, and Game 3 is a dental hygiene educational game, named ToothGame.

In Game 1, the player is the mayor of Green City, and needs to construct different buildings to develop the city. Different kinds of buildings have different effects, such that industrial buildings could improve the economy, but would cause environmental damage. Conversely, parks do not help the economy, but can help preserve the environment. Thus, players must use knowledge to balance between economic development and environmental protection. There are many levels to the game with different task objectives and differing levels of difficulty, When a player completes the initial levels, s/he will receive questions about environmental protection. If the player answers the questions correctly, the next level is unlocked. Through this game, players understand the cost of economic development and learn methods of protecting the environment.

The game describing cardiac catheterization, named National, was developed for the purpose of guiding players toward understanding the cardiac catheterization procedure in order to reduce the anxiety of the player about to undergo a cardiac catheterization. This type of game is a role-playing one, in which players can play the hospitalized patients and experience the process of cardiac catheterization. In addition, there are some questions that the players need to answer in the process of the cardiac catheterization. Players, thus, can reduce the anxiety and restlessness of those required to undergo surgery in this game.

Game 3, named ToothGame, is similar to Whac-A-Mole in that it teaches children the concept of dental hygiene. This game has three levels – children's (easy), youth's (mid-level) and adult's (hard) – which have, correspondingly, different themes and different "right" tools. Players need to choose the right tools and use their finger clicks to eliminate tooth decay within the minute allotted. As the game proceeds, players also need to answer questions about dental health. Total scores are based on the amount of tooth decay eliminated and the number of questions answered correctly.

Participants and evaluation process

Twelve experts, with experience in game learning (more than five years of playing more than ten types of games), were invited to participate in this case study. All of the experts had a basic knowledge of smart phone with an Android OS.

A developed game-evaluation system was used as the main tool in rating educational games. The experts can read the game introduction, play and rate the game, using this system. In order to understand the game, each expert plays each game for a minimum of 10 minutes; there is no maximum time limit. The process of evaluation follows the process described in the system development. Lastly, after the rating process, all of the experts fill out a system satisfaction questionnaire.

Analysis of the learning performance and satisfaction

With random sampling, three classes in three schools in Taiwan were selected for this study. For assessing the learning performance and satisfaction of each of the three types of game-based learning, each class was divided into an experimental group and a control group. For Game 1, 16 students, 8 males and 8 females, were in the experimental group; and 17 students, 9 males and 8 females, in the control group. For Game 2, 21 students, 11 males and 10 females, were in the experimental group; and 22 students, 11 each, male and female, in the control group. For Game 3, 18 students, 9 each, male and female, were in the experimental group; and 19 students, 10 males and 19 females, in the control group. A pretest-posttest nonequivalent-groups design in the quasi-experiment was utilized in this study. The experimental groups were instructed with a Game-based Learning System, while the control groups were taught with paper-based materials. Both groups were presented with the same textbook units and then proceed to the Learning Performance Evaluation Pretest and Posttest. Cronbach's α was applied to calculate the reliability of the test questions and the questionnaires, wherein the level of each dimension was above the standard, 0.7, suggested by Hair et al. (1998), presenting a scale with certain reliability. The learning performance on these three types of games (Table 3) reveals no significant difference in the pre-test, but the experimental groups adduce remarkably larger differences than the control groups. The experimental groups present higher satisfaction, with the three types of game-based learning. This confirms their learning performance and satisfaction with a game-based learning system, and shows that the learning effects have been reinforced.

System satisfaction questionnaire

A system satisfaction questionnaire was used to understand the efficaciousness of the developed evaluation system. There are two standard ways of eliciting feedback: the conventional paper-and-pencil questionnaire, as noted above; and through e-games. Our developed system questionnaire has 8 items by Likert, based on a five-point scale.

Comparison analysis of the system effectiveness

In order to evaluate the system's effectiveness, both an SUS Satisfaction Questionnaire and a System Efficiency Assessment Comparison were utilized. The system satisfaction questionnaire was used to understand user satisfaction, so as to evaluate the system. Meanwhile, traditional assessment methods (arithmetic average), EMGS (with no weight), EMGS (with OWA weight), and an expert assessment method (the Delphi Method) were utilized for assessing the efficiency of the EMGS model. The Delphi Method, as the expert assessment method, was first used as the comparison criteria. The Delphi Method, proposed by Dalkey and Helmer from RAND Corporation in the early 1950s, is a systematic procedure designed to express the consensus of a group of experts. With the questionnaire survey, the experts are repeatedly and anonymously questioned. In succeeding questionnaires, a coordinator returns to the results of the previous questionnaire, so

Table 3. Analysis of game-based learning performance and satisfaction

		Group	Mean	SD	Satisfaction	F
Game 1 (n=33)	**Pretest**	EG(n=16)	78	11	X	0.52 ns
		CG(n=17)	75	10	X	
	Posttest	EG(n=16)	85	8	4.6	27.30***
		CG(n=17)	74	12	3.1	
Game 2 (n=43)	**Pretest**	EG(n=21)	76	12	X	0.42 ns
		CG(n=22)	77	15	x	
	Posttest	EG(n=21)	85	10	4.5	18.40***
		CG(n=22)	76	15	3.4	
Game3 (n=37)	**Pretest**	EG(n=18)	74	12	X	0.34 ns
		CG(n=19)	75	11	X	
	Posttest	EG(n=18)	89	8	4.4	22.10***
		CG(n=19)	77	14	3.4	

***p< 0.001. ns= no significant (CG=Control Group, EG=Experiment Group)

Table 4. Results of the evaluation with α is 0.6

Dimensions	D1	D2	D3	D4	D5	D6	D7	D8	I
Game 1	5.208	4.958	4.979	6.233	5.771	5.667	4.000	5.208	0.501
Game 2	4.917	3.583	4.188	4.117	4.563	3.708	5.583	6.375	0.382
Game 3	7.583	7.667	6.875	7.367	8.125	7.625	3.667	6.375	0.695

Table 5. Comparison and analysis of the system efficiency

	Game 1				Game 2				Game 3			
	M1	M2	M3	M4	M1	M2	M3	M4	M1	M2	M3	M4
D1	6	5	4	4	2	4	3	3	1	3	4	4
D2	5	6	7	8	6	8	8	8	2	4	2	2
D3	7	7	6	6	7	5	5	5	7	5	6	6
D4	3	2	1	1	5	6	6	6	8	6	5	5
D5	4	1	2	2	1	3	4	4	3	2	1	1
D6	2	4	3	3	8	7	7	7	4	1	3	3
D7	8	8	8	7	3	1	2	2	6	8	8	8
D8	1	3	5	5	4	2	1	1	5	7	7	7
Order	3	1	2	2	1	2	3	3	3	2	1	1

M1=Traditional evaluation; M2= EMGS (no weight); M3= EMGS (with weight); M4=Experts (Delphi method)

that the experts can understand their peers' opinions and review their own, until a consistent opinion is reached. The Delphi Method is, therefore, applied to assessing the efficiency of the proposed EMGS. The expert panel was comprised of 11 academic experts and business representatives with particular knowledge, or with more than 10 years of experiences in game design and teaching design. In regard to the system efficiency assessment, different methods generate distinct calculation results. In order to effectively compare various assessment methods, a consistent sorting method was utilized, in which number 1 stands for the most important sorting, while number 8, for the least important. In Table 5, the sorting values of the dimensions in the EMGS (with OWA weight) are consistent with the ones in the experts' assessment method; especially, those in Game 3>Game 1>Game 2, which, approaching the experts' assessment criteria, appears to reveal the accuracy of the proposed EMGS (with OWA weight). Since the EMGS (with OWA weight) is a mobile assessment system, its accuracy better approaches the experts' opinion than would accuracy measured by traditional assessments (M1, M2), and it can more efficiently complete the complicated calculations.

Result and finding

Table 4 shows the result of our case study with the dimension scores of each game. The dimension score is the average from each scored item it contains. Dimension D1 is clear goals, D2 is feedback, D3 is challenge and skills, D4 is sense of control, D5 is concentration, D6 is immersion, D7 is interactivity, and D8 is knowledge improvement. Further, the "I" value is calculated by the OWA & TOPSIS methods. Under the Flow dimension, there are 8 items, under situational values (α), there are 0.6 items, and importance ordering is D4>D5>D1>D3>D2>D6>D8>D7.

Game 1 gains a medium score for each dimension between 4 and 6.233, and its integrated value (I value) is 0.502. Although Game 2 is the worst alternative, with an I value of 0.385 and the lowest scores in dimensions 1, 2, 3, 4, 5 and 6, it has the highest score in dimensions 7 and 8. Because 2 is a role-playing game, the player plays a patient, talking to nurses or doctors in the hospital. This kind of game is best suited for an education in which the learners can learn through conversations with virtual characters. Thus, game 2 also gained the highest score on interactivity and knowledge improvement. There is no doubt game 3 is the best educational game with the highest I value of 0.697, the highest scores in dimensions 1, 2, 3, 4, 5 and 6, and with a dimension 8 score the same as game 2. Game 3 is similar to Whac-A-Mole with a clear goal and easy gameplay. Players only need to click their fingers to eliminate tooth decay and obtain points. The score appears on the screen, which enables players to understand their own skill level. Among several different challenges of varying skill levels, a player can select the one which best matches his/her own skill set. This game asks the same dental hygiene question to each player and provides the correct answer if the player answers incorrectly. It has numerous advantages as an educational game, such as providing a clear goal, feedback and a sense of control etc.; also, it quickly grasps the player's attention, leading him/her to a Flow state. In summation, of the three educational games, game 1 has the most advantages, but its interactivity (Dimensions 7) is the lowest; game 2 has an average score and no obvious shortcomings; and, finally, although, game 3 has the lowest score in most dimensions, having the highest score in interactivity makes it the best alternative for interactive teaching.

The weight of Flow dimension calculated by the OWA method with a different α is shown in Table 6. Chang, Cheng & Chen (2006) confirmed that the result is equal to $\alpha = 0.5 + \delta$ and $\alpha = 0.5 - \delta$（$0 \leqq \delta \leqq 0.5$）, thus, only showing weights with an α greater than 0.5. W_1 is the weight of the most important dimensions, W_2 is the weight of the second most important, W_3 is the weight of the third most important, etc.

Table 6. Weight of Flow dimension with α = 0.5 to 1

Weight	α=0.5	α=0.6	α=0.7	α=0.8	α=0.9	α=1
W_1	0.12500	0.19175	0.27925	0.39921	0.58635	1
W_2	0.12500	0.16737	0.20892	0.24283	0.24284	0
W_3	0.12500	0.14609	0.15631	0.14771	0.10057	0
W_4	0.12500	0.12752	0.11695	0.08984	0.04165	0
W_5	0.12500	0.11130	0.08749	0.05465	0.01725	0
W_6	0.12500	0.09715	0.06546	0.03324	0.00714	0
W_7	0.12500	0.08480	0.04898	0.02022	0.00296	0
W_8	0.12500	0.07402	0.03664	0.01230	0.00123	0

Table 7. I value of each game with all α (importance ordering: 4 5 1 3 2 6 8 7)

	α=0.5	α=0.6	α=0.7	α=0.8	α=0.9	α=1
Game 1	0.469	0.501	0.529	0.551	0.570	0.581
Game 2	0.407	0.382	0.369	0.361	0.354	0.346
Game 3	0.649	0.695	0.717	0.721	0.717	0.707
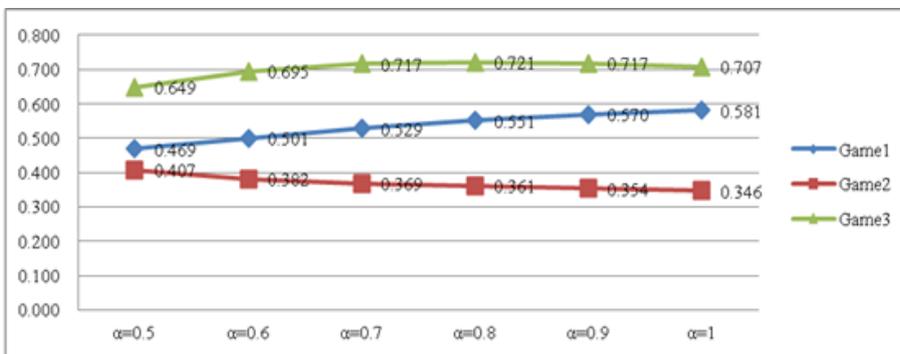
Figure 7. I value of each game with all α (importance ordering: 4 5 1 3 2 6 8 7)

Table 7 and Figure 7 show that each game gains I value in each α value, and that the importance ordering is D4>D5>D1>D3>D2>D6>D8>D7. The results show that a different α value did not change their ranking.

Finally, the score is the average of all the questionnaires. All items scored more than 4 points, with SS8 having the highest scores of 4.95, and SS4 having the lowest rating with 4.08.

Finding 1: Flow ingredient

In this study, Flow dimension is divided into eight ingredients: clear goal, feedback, challenge and skills, sense of control, concentration, immersion, interactivity, and knowledge improvement. Generally, the more Flow dimensions a mobile-learning game has, the greater the probability of Flow state occurrence. Specifically, a mobile-learning game with medium scores in all Flow dimensions is better than a game with high scores in a few dimensions and low scores in numerous dimensions, even if both sets of scores have the same integrated value I. This is because an extremely low Flow dimension score diminishes player attention and undermines the Flow state, even if it only has one Flow dimension. Thus, game designers must ensure that each dimension has a basic level of Flow dimension.

Although a qualified mobile-learning game should have a basic level of all Flow dimensions, there is still a greater influence from particular Flow dimensions. Ordering importance will be discussed in finding 3.

Finding 2: Conservative situation

Situational parameter (α) refers to the view of decision makers for the evaluation weight in the OWA method. Generally, an α value approaching 1 means that the decision maker is optimistic, which results depend on a several factors. In contrast, an α value closer to 0.5 represents pessimism, because the evaluation results reference all factors.

Table 6 shows each weight of Flow dimension with an α value of 0.5 to 1. This demonstrates that, with an α closer to 1, the weight of the dimension becomes greater, and appears among the top few in the importance ordering. Specifically, the top few Flow dimensions in the importance ordering decide the ranking result under the higher α value, so that the game obtains a high integrated value I by having only a few dimensions with high scores. However, a game with Flow cannot dependent on a few Flow dimensions to enhance the Flow state, because adverse dimensions would undermine it. Furthermore, all Flow dimensions have equal weight when the α value is 0.5, which is not conducive to evaluating games with Flow, because the importance of the dimensions has not been given consideration. For the above reasons, this study recommends that the value of α should approach 0.6 before considering the sufficiency and importance of the flow dimension.

Finding 3: Preference ordering

This study identified Flow dimension ordering importance through discussion among experts, based on mobile-learning games. The ordering is D4 (sense of control), D5 (concentration), D1 (clear goals), D3 (challenge and skills), D2 (feedback), D6 (immersion), D8 (knowledge improvement), and D7 (interactivity). Educational games should make players feel a sense of control in both the interface and the gameplay. This means that the interface should be simple and the gameplay easy to understand. How the player feels about using the interface will greatly affect whether s/he will continue playing the game. Secondly, the game must provide a lot of stimuli from different elements in order to quickly grab the player's attention, and the game tasks should be related to learning objectives. Clear goals will not confuse the player, and they will also keep the player's attention focused on the task. Finally, the game should provide an appropriate challenge to avoid having players feel anxious or bored.

In fact, different importance ordering could lead to different results. The importance ordering is determined by e-learning educators who teach interactively. Thus, the D7 (interactivity) is the most important, D8 (knowledge improvement) is of secondary importance, while the others remain unchanged. The results of a different importance ordering are shown in Table 8 and Figure 8, which is completely different from the results of the original importance ordering (D4>D5>D1>D3>D2>D6>D8>D7). In $\alpha = 0.6$, the Integrated value - I value of game 2 is higher than that of game 1, therefore, it becomes the best alternative in $\alpha = 0.8$.

In sum, there are different importance orderings in different situations, such as teaching methods and course requirements, and these will lead to different rank results. Thus, allowing educators to adjust the importance ordering, according to their preferences, contributes to ranking education games more flexibly and with more accurate results.

Table 8. I value of each game with all of α (importance ordering: 7 8 4 5 1 3 2 6)

	α = 0.5	α = 0.6	α = 0.7	α = 0.8	α = 0.9	α = 1
Game 1	0.469	0.448	0.416	0.377	0.333	0.296
Game 2	0.407	0.453	0.488	0.510	0.517	0.509
Game 3	0.647	0.583	0.515	0.450	0.386	0.333
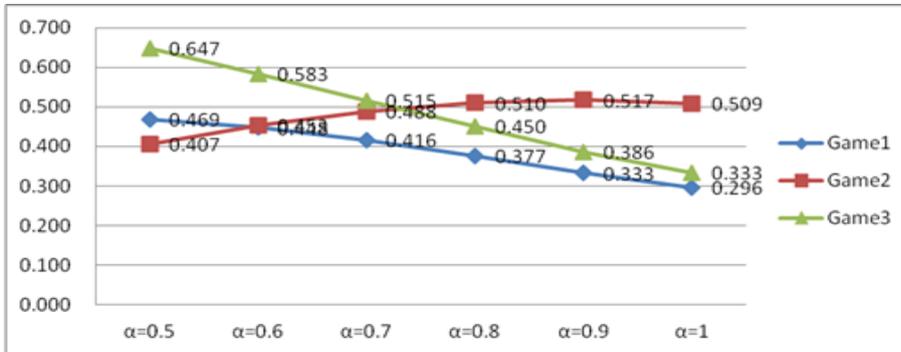
Figure 8. I value of each game with all of α (importance ordering: 7 8 4 5 1 3 2 6)

Finding 4: System satisfaction

The developed EMGS provides a process, including game description, play game, game rating and a calculated ranking result. In which all items gained 4 points or more, shows an overall satisfaction with the system. The lowest score, 4.08 on SS4, may be due to some raters being accustomed to paper-and-pencil questionnaires. SS8 achieved the highest score, 4.95, which, relative to paperwork usage, shows the rank result immediately.

In summary, the evaluation process with a developed system provides better results than the conventional, paper-and-pencil method. Using a developed system to evaluate games is faster and more convenient, regardless of whether the comparison is with a conventional process, and rating each item of each game can effectively help with the comparisons of the game differences. Finally, the biggest difference between using our developed system and a conventional process is that the rater can immediately ascertain results. The results of the system satisfaction questionnaire prove that our evaluation method and our system are, indeed, both straightforward and quick.

CONCLUSION

A novel MCDM model, based on Flow theory, has been proposed to evaluate mobile-learning games for assisting educators in selecting the best educational games. In order to evaluate educational games, this study identified eight Flow dimensions with 25 items noted previously in the literature and in the Delphi method. The importance ordering of the Flow dimension is given by experts, based on the situation of the educator, in descending order as: sense of control, concentration, clear goal, challenge and skills, feedback, immersion, knowledge improvement and interactivity. This study developed an EMGS evaluation system, based on Flow dimension, which provides an evaluation process, including viewing the game description, playing the game, rating the game and viewing the ranking result. The OWA & TOPSIS method was employed to rank games and calculate the weight of the Flow dimension by the OWA method, with an initial situational parameter of α as 0.6. However, educators can adjust the importance ordering, according to their individual situations, such as teaching methods and course requirements, which would allow more flexibility and accuracy in the rank result.

The case study and system satisfaction questionnaire verified that the evaluation process provided by this system is fast and convenient, and that it can give an accurate recommendation to assist educators in selecting the most appropriate games for learning.

Future research will build a list of Flow dimensions for evaluating games, which will allow raters to choose among them, based on their own preferences and priorities, and will cover more types of games to be evaluated by our system, e.g., console games. Finally, we will use other rank algorithms, such as the fuzzy TOPSIS method, to rank games in order to improve the accuracy of the rank result.

REFERENCES

Adler, M. (1996). Gazing into the oracle: the Delphi method and its application to social policy and public health. Jessica Kingsley Publishers.

Agarwal, R., & Karahanna, E. (2000). Time flies when you're having fun: cognitive absorption and beliefs about information technology usage. MIS Quarterly, *24*(4), 665-694.

Anderson, D. R., Sweeney, D. J., & Williams, T. A. (1998). Quantitative methods for business (7th ed.). OH: International Thomson Publishing.

Behzadian, M., Otaghsara, S. K., Yazdani, M., & Ignatius, J. (2012). A state-of-the-art survey of TOPSIS applications. *Expert Systems with Applications*, *39*(17), 13051-13069.

Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: a theoretical synthesis. *Review of Educational Research*, *65*(3), 245-281.

Chang, J. R., Ho, T. H., Cheng, C. H., & Chen, A. P. (2006). Dynamic fuzzy OWA model for group multiple-criteria decision making. *Soft Computing*, *10*(7), 543-554.

Chen, H., Wigand, T. R., & Nilan, M. (1999). Optimal experience of web activities. *Computers in Human Behavior*, *15*(5), 585-608.

Cheng, C. H., & Lin, Y. (2002). Evaluating the best main battle tank using fuzzy decision theory with linguistic criteria evaluation. *European Journal of Operational Research*, *142*, 174-186.

Choi, D. H., Kim, J., & Kim, S. H. (2007). ERP training with a web-based electronic learning system: the Flow theory perspective. *International Journal of Human-Computer Studies*, *65*(3), 223-243.

Clayton, M. J. (1997). Delphi: a technique to harness expert opinion for critical decision making tasks in education. *Educational Psychology*, *17*(4), 373-386.

Csikszentmihalyi, M. (1975). Beyond boredom and anxiety. San Francisco: Jossey-Bass.

Csikszentmihalyi, M., & Kleiber, D. A. (1991). Leisure and self-actualization. In B. L. Driver, P. J. Brown & G. L. Peterson (Eds.), Benefits of Leisure (pp.91-02). State College, PA: Venture.

C.Y. Mao, Q. Mei, Z. Ma.(2009).A new method for information system selection, in: Second International Conference on Future Information Technology and Management Engineering, IEEE, http://dx.doi.org/10.1109/FITME.2009.22

Dickey, M. D. (2005). Engaging by design: how engagement strategies in popular computer and video games can inform instructional design. *Educational Technology Research and Development*, *53*(2), 67–83.

Dondi, C., & Moretti, M. (2007). A methodological proposal for learning games selection and quality assessment. *British Journal of Educational Technology*, *38*(3), 502–512.

Eow, Y. L., Ali W. Z. W., Mahmud, R., & Baki, R. (2009). Form one students' engagement with computer games and its effect on their academic achievement in a Malaysian secondary school. *Computers & Education*, *53*(4), 1082-1091. Doi: 10.1016/j.compedu.2009.05.013.

Finneran, C. M., & Zhang, P. (2003). A person-artifact-task (PAT) model of Flow antecedents in computer-mediated environments. *International Journal of Human-Computer Studies*, *59*(4), 475-496.

Fish, L. S., & Busby, D. M. (1996). The Delphi method. In D. H. Prenkl & S. M. Moon (Eds.), Research methods in family therapy (pp. 469–482). NY: Guilford Press.

Fu, F. L., Su, R. C., & Yu, S. C. (2009). EGameFlow: a scale to measure learners' enjoyment of e-learning games. *Computers & Education*, *52*(1), 101-112.

Fullér, R., & Majlender, P. (2001). An analytic approach for obtaining maximal entropy OWA operator weights. *Fuzzy Sets and Systems*, *124*(1), 53-57.

García-Cascales, M. S., & Lamata, M. T. (2012). On rank reversal and TOPSIS method. *Mathematical and Computer Modelling*, *56*(5), 123-132.

Goodman, C. (1987). The Delphi technique: a critique. *Journal of Advanced Nursing*, *12*, 729-734.

Hoffman, D. L., & Novak, T. P. (2000). *The growing digital divide: implications for an open research agenda*. Understanding the Digital Economy. MIT Press, Cambridge, MA.

Hoffman, D. L., & Novak, T. P. (2009). Flow online: lessons learned and future prospects. *Journal of Interactive Marketing*, *23*(1), 23-34.

Hogle, J. G. (1996). Considering games as cognitive tools: in search of effective edutainment. Retreived November 15, 2011, from http://twinpinefarm.com/pdfs/games.pdf.

Hoffman, D. L., & Novak, T. P. (1996). Marketing in hypermedia computer-mediated environments: conceptual foundations. *The Journal of Marketing*, 50-68.

Hsieh, L. F., Chin, J. B., & Wu, M. C. (2006). Performance evaluation for university electronic libraries in Taiwan. *The Electronic Library, 24*(2), 212-224.

Hsu, C. L., & Lu, H. P. (2003). Why do people play on-line games? An extended TAM with social influences and Flow experience. *Information & Management, 41*(7), 853-868.

Huang, M. H. (2003). Designing website attributes to induce experiential encounters. *Computers in Human Behavior, 19*(4), 425-442.

Hwang, C. L. & Yoon, K. (1981). Multiple Attribute Decision Making: methods and applications. New York: Springer-Verlag.

Jee, D.H., & Kang, J.K. (2000). A method for optimal material selection aided with decision making theory. *Materials and Design, 21*(3), 199-206.

Johnson, L., Smith, R., Willis, H., Levine, A., & Haywood, K. (2011). The 2011 Horizon Report. Austin, Texas: The New Media Consortium.

Kebritchi, M., Hirumi, A., & Bai, H. (2010). The effects of modern mathematics computer games on mathematics achievement and class motivation. *Computers & Education, 55*(2), 427-443.

Kiili, K., de Freitas, S., Arnab, S., & Lainema, T. (2012). The design principles for Flow experience in educational games. *Procedia Computer Science, 15*, 78-91.

Kua-Hsin Peng, Gwo-Hshiung Tzeng (2013), A hybrid dynamic MADM model for problems-improvement in economics and business, *Technological and Economic Development of Economy, 19*(4), 638-660

Li, S. (2005). A Web-enabled hybrid approach to strategic marketing planning: group Delphi + a Web-based expert system. *Expert Systems with Applications, 29*, 393-400.

Linstone, H. A., & Turoff, M. (Eds.). (1975). The Delphi method: techniques and applications (pp. 37-71). Boston, MA: Addison-Wesley Publishing Company, Advanced Book Program.

L. Blanc, M. Jelassi.(1989).DSS software selection: a multiple criteria decision methodology, *Inform. Manage, 17*,49–65.

McFarlane, A., Sparrowhawk, A. & Heald, Y. (2002). Report on the educational use of games: an exploration by TEEM of the contribution which games can make to the education process. (On line). Available: http://reservoir.cent.uji.es/canals/octeto/es/440.

Murry Jr, J. W., & Hammons, J. O. (1995). Delphi: a versatile methodology for conducting qualitative research. *Review of Higher Education, 18*(4), 423-36.

M. Morisio, A. Tsoukis. (1997) .IusWare: a methodology for the evaluation and selection of software products, *IEEE Proc. Softw. Eng. 144* (3),162–174.

Novak, T. P., Hoffman, D. L., & Yung, Y. F. (2000). Measuring the customer experience in online environments a structural modeling approach. *Marketing Science, 19*(1), 22-42.

Oblinger, D. G. (2004). The next generation of educational engagement. *Journal of Interactive Media in Education, 2004*(8), 1-18.

O'Hagan, D. T., Palin, K. J., & Davis, S. S. (1988). Intestinal absorption of proteins and macromolecules and the immunological response. *Critical Reviews in Therapeutic Drug Carrier Systems, 4*(3), 197.

Prensky, M. (2001). Digital game-based learning. New York: McGraw-Hill.

Provenzo, E.F. (1991). *Video kids: making sense of Nintendo.* Cambridge, MA: Harvard University Press

Reid, N. (1988). The Delphi technique: its contribution to the evaluation of professional practice. In R. Ellis (Ed.), *Professional competence and quality assurance the caring professions* (pp. 230–262). New York: Chapman & Hall.

Saaty, T. L. (1980). *The analytic hierarchy process*. McGraw-Hill, New York.

Salen, K., & Zimmerman, E. (2004). Rules of play (ROP): game design fundamentals. Cambridge: MIT Press.

Shaffer, D, W. (2006). Epistemic frames for epistemic games. *Computers & Education, 46* (3), 223-234.

Shih, H.S, Syur, H.J, & Lee, E.S. (2007). An extension of TOPSIS for group decision making. *Mathematical and Computer Modeling, 45*, 801-813.

Skadberg, Y. X., & Kimmel, J. R. (2004). Visitors' Flow experience while browsing a web site: its measurement, contributing factors and consequences. *Computers in Human Behavior, 20*, 403-422.

Sweetser, P., & Wyeth, P. (2005). GameFlow: a model for evaluating player enjoyment in games. *ACM Computer in Entertainment, 3*(3), 1-24.

Tiger, L. (2000). The pursuit of pleasure. Boston: Little, Brown and Company. pp. 52-60.

Turban, E., & Aronson, J. (2001). Decision support systems and intelligent systems (6th ed.). Upper Saddle River, NJ: Prentice-Hall.

Webler, T., Levine, D., Rakel, H., & Renn, O. (1991). A novel approach to reducing uncertainty: the group Delphi. *Technological Forecasting and Social Change, 39*(3), 253-263.

Yager, R. R. (1988). On ordered weighted averaging aggregation operators in multicriteria decision making. Systems, Man and Cybernetics, IEEE Transactions on, 18(1), 183-190.

Yager, R. R. (2004). On the retranslation process in Zadeh's paradigm of computing with words. Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on, 34(2), 1184-1195.

Yong, D. (2006). Plant location selection based on fuzzy TOPSIS. *International Journal of Advanced Manufacturing Technology, 28*(7-8), 839-844.

Zanakis, S.H., Solomon, A., Wishart, N., & Dublish, S. (1998). Multi-attribute decision making:A simulation comparison of select methods. *European Journal of Operational Research, 107*(3), 507-529.

◈ ◇ ◈