

Goal Question Metric as an Interdisciplinary Tool for Assessing Mobile Learning Application

Sim Yee Wai¹, Cheah WaiShiang², Piau Phang³, Kai-Chee Lam⁴, Eaqerzilla Phang⁵, Nurfaeza binti Jali⁶

Faculty of Computing and Engineering, Quest International University (QIU), Perak, Malaysia¹

Faculty of Language and Communication, Universiti Malaysia Sarawak, Kota Samarahan, Malaysia⁴

Faculty of Computer Science & IT, Universiti Malaysia Sarawak, Kota Samarahan, Malaysia^{2,3,5,6}

Abstract—Assessing the mobile learning application among interdisciplinary researchers is a non-trivial task. Mandarin Learning App is a Mandarin 3D game tailor-made for students who choose PBC1033 Mandarin Language Level 1 as an elective course. It is an interdisciplinary project which it involves researchers from software engineering, computational science/mathematics and the Faculty of Language studies. In the project, the software engineer focuses on producing a quality application mostly through usability studies; the language teacher focuses on students' study performance upon using the Mandarin Learning App and the mathematician focuses on finding the statistical data dependency of collected data through the various statistical packages. Hence, we are facing issues like how to reach a consensus in working on assessing the Mandarin 3D games? How to enable the discussion among the researchers; how to consolidate the results so that we can understand? We introduce Goal Question Metric to tackle these issues. In this paper, we demonstrate how Goal Question Metric is used to form a holistic view of assessing requirements on mobile applications and guide the discussion and reach consensus in analyzing the results of the evaluation. The contribution of this paper is to introduce Goal Question Metric as an interdisciplinary tool while assessing the mobile learning application. With Goal Question Metric, we demonstrate how it can structure the assessment from a different viewpoint in a comprehensive and systematic manner; 1) better structure of the experiments, 2) able to reach consensus among researchers from different disciplines, 3) able to analyze the dependencies among various experiments and 4) able to find hidden results.

Keywords—Goal question metric; mobile application; evaluation; communication; interdisciplinary

I. INTRODUCTION

Language is very important for everyone to have connection and interaction with others either in verbal or nonverbal communication. UNIMAS has offered introductory Mandarin Language courses for students who wish to explore and learn Mandarin. Among them, the PBC0033 Mandarin Language Level 1 is offered as an elective course to students who do not have any basic knowledge of Mandarin. In this course, students start to learn and recognize the Chinese character; to pronounce the character together with pinyin and to write the character with the correct stroke. Based on our informal observation, students are interested to adopt mobile applications to learn Mandarin. Hence, we developed an in-house mobile application known as "Mandarin App", under an interdisciplinary project between researchers from software engineering, language teacher and mathematician.

In the project, the software engineer focuses on producing a quality application by evaluating the system mostly through usability studies. On the other hand, the language teacher focuses on students' study performance upon using the Mandarin Learning App and the mathematician focuses on finding the statistical dependency of collected data through the various statistical packages. Hence, we are facing issues like how to reach a consensus in working on assessing the Mandarin 3D games? How to enable the discussion among the researchers? How to understand the assessing mechanism in which some are too technical (e.g., statistical analysis) and how to consolidate the results so that we can understand? Hence, we are in the dilemma to reach a consensus when conducting the evaluation of the project.

We introduce Goal Question Metric (GQM) to tackle these issues. In this paper, we demonstrate how Goal Question Metric is used to form a holistic view of assessing requirements on mobile applications and guide the discussion and reach consensus in analyzing the results of the evaluation. The contribution of this paper is to introduce GQM as an interdisciplinary tool while assessing the mobile learning application. With GQM, we demonstrate how it can structure the assessment from a different viewpoint in a comprehensive and systematic manner. Section II presents the related works in using GQM to assess a system. Section III presents how GQM is used to assess a mobile application. Section IV presents the results derived from GQM measurement guidelines. The paper is concluded in Section V.

II. RELATED WORK

GQM is used to perform usability evaluation of real-time water quality monitoring mobile applications [1] and general mobile applications [2], [3], [4], [5], [6], [7], [8], [9]. It describes the data that we need to collect and how to interpret the data [10]. The GQM consists of goal, questions, and metrics. Several metrics are identified in user experiment measurements on mobile applications [11]. They are understandability, learnability, efficiency, effectiveness, operability, attractiveness, usability compliance, happiness, engagement, adoption, retention, task success, usability, satisfaction in use, safety, social presence, cross-platform interaction, algorithm diversity, user control, usage effort, outcome-related experience, review, connection, content, internet service, service availability, ease to use, utility, long term use, productivity, generalizability, pragmatic quality, hedonic quality, stimulation, emotion, psychology metric.

Other metrics are task completion, error rates, time to take the usage and subjective satisfaction like enjoyment, ease of use and safety [1]. Saleh et al. [12] classified the usability metrics into 7 criteria. They are satisfaction, efficiency, effectiveness, learnability, operability, universality, and user interface aesthetics. The satisfaction covers metrics of comfort, trust, pleasure, usefulness; the efficiency covers metrics like task efficiency, time efficiency, relative task time; effectiveness covers metrics of task completion, task effectiveness and effort frequency; the learnability covers metrics of time to learn, memorability, easy to understand error messages, completeness of user documentation, cognitive load; operability covers metrics of understandable, message clarify, operational consistency; universality covers metrics of cultural universality, standard compliance, accessibility; user interface aesthetics covers metric of customizability and attractive of the user interface.

The GQM is extensible to serve as a measurement guideline to measure mobile applications [1]. For example, [1] extend the GQM with guidelines to determine how a specific goal is reached. The guideline can drive the formation of questions for GQM. Meanwhile, the guideline will serve as expected answers for the metrics. For example, to achieve goal simplicity, a system should be easy to input the data, easy to install, easy to learn. Hence, the evaluator can form the questions like is it simple to key in data? How easy it is to install the application? etc. In addition, GQM [10] has been used in running a course activity to train the student's ability to understand its goal, refine its goal and design appropriate achievement metrics.

We extend the usage of GQM from usability evaluation into a wider assessment mechanism. Hence, GQM for mobile application evaluation should not only focus on usability study of mobile applications but it can also be used and support various experiments analysis. This is in line with the work [1] to extend the usage of GQM with guidelines. The guideline presents the expected answers for the metrics. On the other hand, [13] extended the GQM with instrument types and role mapping to relate the evaluation and analysis methods into metrics. Meanwhile, our work is related to [10] by treating GQM as a tool for stakeholders during the assessment of mobile learning application.

III. THE METHOD

This section presents the adoption of GQM as an interdisciplinary tool to assess mandarin learning app among different researchers.

GQM is known as Goal, Question, Metric [14]. The GQM model is represented as a hierarchical structure and covers three levels. They are conceptual level, operation level and quantitative level. The conceptual level defines the goal of the measurement. It represents the measurement from a high level of abstraction. The goal at this level covers the quality level of the system. Hereby, we treat is quality goal. The quality goal is to act as a non-functional goal and represents how a functional goal should have been achieved. The quality goal is the property of the system. Quality goals are non-functional goals (sometimes referred to as soft goals). The operation level defines a set of questions that will work towards the

assessment or achievement of a specific goal. The quantitative level defines a set of data that serves as the answers to the questions. The data can be objective or subjective depending on the measurement mechanisms (e.g. quantitative or qualitative assessment method).

The GQM processes cover the process to determine what to measure and determine how to measure. To determine what to measure, it involves processes to identifying entities, classifying entities to be examined and determine the relevant goals. Meanwhile, it is important to inquire about metrics and assign the metrics to determine how to measure.

In this section, we introduce how to adopt GQM as an interdisciplinary tool to assess mandarin learning apps among researchers from different disciplines.

The goal is the first step when working on the GQM model. The goal represents the objective or purpose of a system, and it is achievable [15]. Since GQM is used to measure the system, the goal must relate to what is the purposes to measure the entities regarding the issues been solved by whom. The goal is related to software quality characteristics as described in [1]. The goal here also refers to what do you want to know or learn from the system. The formation of the goal is based on the template given in [16]. According to [16], a goal is derived based on the elements like purpose, issues, object and viewpoint [16]. Goals need to be high level. We can derive the goals by referring to the documents' analysis, interviews [14] or existing usability models as listed in [1], [13], [12], [2], [17].

Questions are elicitation questions in achieving the goal as stated previously. The generation of the questions is based on the instrument decided and targets users.

Metrics are the measurement parameters in answering the questions towards achieving the goal. Metrics produces data. The data can be objective or subjective depending on the types of measurement mechanisms. For example, a quantitative analysis leads to objective data (e.g. number of satisfaction; the number of scores). Meanwhile, the qualitative analysis led to subjective data. Once, the GQM is formed, a goal-question-metrics refinement process takes place to enable the researchers to reformulate the goal, questions, and metrics on evaluating the mandarin learning app.

As mentioned before, this is an interdisciplinary project. The software engineer has the intention to develop a quality product and usability studies are among the famous technique for assessing it. On the other hand, language teacher is always focuses on study performance. This is done through pre-test and post-test. In addition, an empirical study is used to study the correlation of results regarding demographic and various post-survey as proposed by the mathematician. Hence, how to consolidate the different types of analysis and studies?

To evaluate the effectiveness of the games, several experiments are conducted. Data are collected by using both the written questionnaire surveys and interviews, a vocabulary pre-test and post-test (before-after), as well as game diaries. Collected data were analyzed with quantitative and qualitative methods. Specifically, exploratory factor analysis was performed to reexamine the grouping of variables in

questionnaire items to establish underlying dimensions that could explain its correlations.

All the experiments are conducted among 33 students from PBC0033, the batch year 2021 who answered the questionnaire and took the pre-test and post-test. As our sample size is very small, we only remove 3 respondents who rate their responses on Likert-type questionnaire items with the same answer always, resulting in a standard deviation lower than or equal to 0.385. Hence, the characteristics and demographics of our remaining 30 respondents are summarized in Table I.

TABLE I. RESPONDENT DEMOGRAPHICS AND CHARACTERISTICS

Demographic or characteristic	Frequency	Per cent
Gender:		
Male	5	16.5
Female	25	83.3
Ethnic:		
Bumiputera	28	93.3
Dusun	2	6.7
Age:		
21 - 22	20	66.7
23 - 24	8	26.7
25 - 26	2	0.6

Prior to the evaluation, the GQM is adopted and serve as a tool to drive the discussion among the researchers. Based on GQM, Table II shows the list of goals, questions, and metrics in related to our study. Meanwhile, Fig. 1 shows the GQM model in assessing the mobile learning App. It starts with a higher-level goal to ensure the effectiveness of learning through the mobile Mandarin App.

To achieve the main goal, there exists sub-goals of 'ensure engaging', 'ensure high performance', 'ensure likelihood', 'ensure secure learning', 'ensure privacy protection' and 'ensure cross culture learning'. The goals are derived from the Table I. To achieve the subgoals, several questions are generated and the answers from the questions will lead to the achievement of the sub-goals. In this case, the answer from "Do the students like to use the app to learn" will achieve the goal to ensure engaging; the answers from "are the games able to improve students' results" will achieve the goal of ensuring high performance; the answer from 'what are the factors that influence the adoption of games' will achieve three subgoals to ensure likelihood of the learning. In addition, we believe that the answers also lead to achievement of subgoals of namely ensure secure learning, ensure privacy protection, ensure cross culture learning after postmortem. The metrics that are corresponded to the questions are level of preference, level of self-learning, level of games challenges, level of critical thinking; number of pre-test level, number of post-test level, comparison of the results; number of dimensions, list of constructs; number of incidents and amount of lost. Finally, we have mapped the metrics to the testing instruments and the researchers that are handled or initiated by the researchers. The mapping of the metrics into instruments is presented in [13].

In sum, the GQM model can serve as a communication media to discuss among the researchers. Before this, we only identified three goals for the assessment. They are ensuring engaging, ensure high performance and ensure likelihood.

After the postmortem, three subgoals have been identified and three questions are derived. They are ensuring secure learning, ensure privacy protection and ensure cross culture learning although it does not cover in this evaluation. It shows how GQM model can be used to find the hidden evaluation elements during the discussion. In addition, the corresponding questions are 'Does the games secure?' 'Do the games able to protect privacy?' and 'Does the games able to promote cross culture learning?'

TABLE II. THE GQM LIST FOR THE STUDY

GQM		Item
Main Goal	Purpose	Improve
	Issue	Learning of mandarin through
	Object (process) Viewpoint	Mobile learning app Among students
Question		Does it an effective way to use mobile learning app?
Metric		Refer to the rest of the description.
Goal 1	Purpose	Ensure
	Issue	engaging in
	Object (process) Viewpoint	learning through mobile app among students
Question		Do the students like to use the app to learn Mandarin?
Metric		Level of preferences to use mobile games Level of self-learning Level of games challenges Level of critical thinking
Goal2	Purpose	Ensure
	Issue	high result performance
	Object (process) Viewpoint	by learning through mobile app among students
Question		Are the games able to improve student results?
Metric		Number of pre-test level Number of post-test level Comparison of the results
Goal3	Purpose	Ensure
	Issue	likelihood of student behaviour
	Object (process) Viewpoint	by learning through mobile app among students
Question		What are the factors that influences the adoption of games among students?
Metric		Number of grouping/ numbers of dimension List of constructs

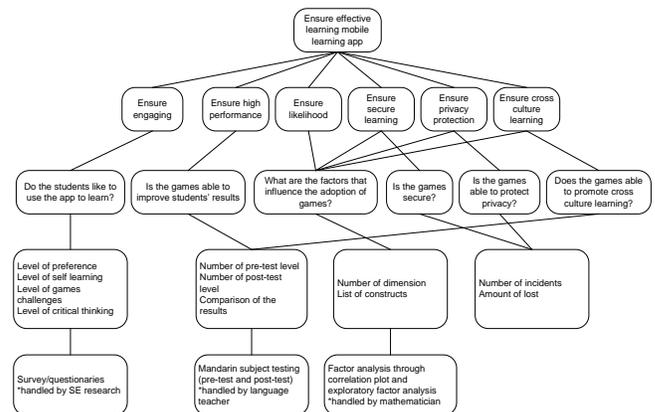


Fig. 1. GQM Model for the Project's Evaluation.

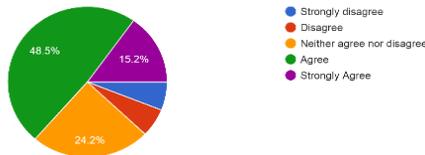
IV. THE RESULTS

A. Results for Goal 1

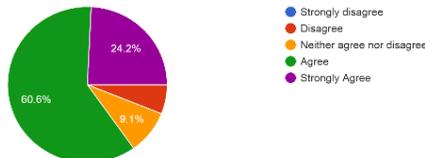
- Question: Is the games able to become a complemented tool on conventional Mandarin lesson?
- Answer: From the results, the mandarin learning app can serve as a complemented tool on conventional mandarin lesson.

Questionnaires are used as the method to achieve the goal. We have formulated the questionnaire based on the metrics given. The results from the usability study are shown in Fig. 2. The survey is designed based on five elements. They are students' motivation; students' attitudes with the games; students' cognitive development; games interface; students' expectation.

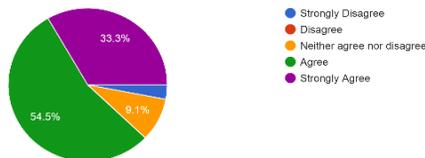
2. I prefer to answer questions using mobile games as compared to using books or paper. 33 responses



3. It is more flexible for me to determine my own learning time. 33 responses



5. These games challenge my understanding of the subject. 33 responses



1. These computer games help me to think critically. 33 responses

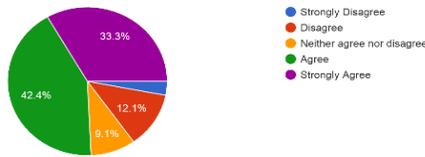


Fig. 2. Usability Study Results.

In sum, students are motivated to use the games in learning Mandarin. 65% of the students are interested to use the games in learning Mandarin instead of using books and paper. The games increase the student learning in which students can learn anytime and anyway (85%). The games can challenge the students to think critically and enhance problem solving in which 76% of respondents stated. Meanwhile, 87% of the

respondents agreed that the games can challenge the students' understanding of Mandarin. It has been reported that the games are interactive and interesting. 79% of the respondents agreed that the menus in the games are easy to understand. 64% of the respondents stated that the navigation and interaction are easy to use. The games are easy to function with a short time learning curve. 91% of the respondents stated that the multimedia elements in the games are interesting. 82% of the respondents show the interest to replicate the games mode in other level of Mandarin course. Although games are interesting, 50% of the respondents still prefer to have the traditional face to face class. This serves the objective on having the games as a complemented tool on conventional Mandarin lesson.

B. Results for Goal 2

- Question: Is the games able to improve student results?
- Answer: The post test results are more than pre- test results. Hence, the games can improve student results.

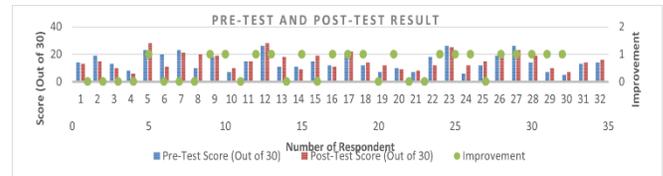


Fig. 3. Individual Pre-test and Post-test Score.

Pre-test and post-test are adopted in this study. We first give the individual pre-test and post-test scores in Fig. 3. Only slightly more than half of the students showed improvement in their test scores after going through mobile games enhanced learning. We then check the normality of the pre- and post-test scores by using Shapiro-Wilk test. Both the test scores are normally distributed as their p-values > 0.05. From their boxplots (see Fig. 4), we find that there is only a marginal improvement whereby the means of pre-test and post-test score are 14.4 and 15.4, respectively. Although the paired sample t-test supports that the true difference in means between pre- and post-test score is not equal to 0, if we plot the change scores (i.e. post-test minus pre-test scores) against pre-test scores, as shown in Fig. 5, a negative correlation can be observed. This implies that a common statistical phenomenon in repeated measurements, known as regression towards the mean (RTM) [20], can influence the findings from our pre- and post-test instruments. This RTM suggests that students with higher pre-test scores consistently make smaller improvement in post-test than students with lower pre-test scores.

As we are not able to use a randomized control group to deal with this RTM, we opt to assess the agreement between these two pre- and post-test measurements through Bland-Altman plot (see Fig. 6). The black line gives the mean of difference in score between pre- and post-test while the two blue dashed lines represents the 95% confidence interval limits of agreement for the mean of difference. In our example, the mean of difference is 0.96. This suggests that on average the post-test measures 0.96 score more than the pre-test as mean of difference (called bias) is non-zero. Besides, only two points appear to lie outside the limits of agreement indicating there is certain degree of agreement between the two tests.

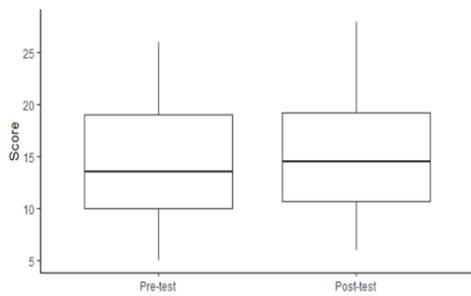


Fig. 4. The Boxplots for Pre- and Post-test Score.

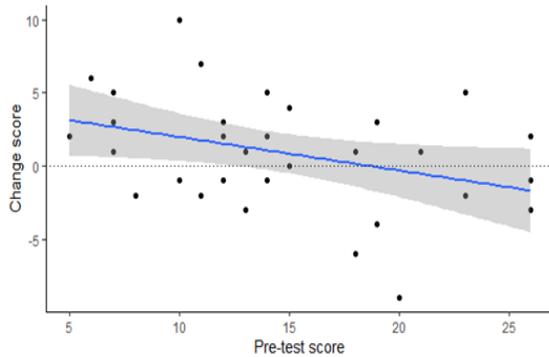


Fig. 5. Scatter Plot showing Change Score against Pre-test Score.

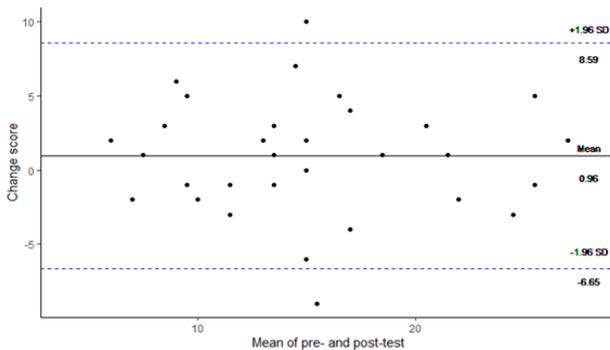


Fig. 6. Bland-Altman plot with the Representation of the Limits of Agreement (Dotted Line).

C. Results for Goal 3

- Question: What are the factors that influence the adoption of games among students?
- Answer: They are two major dimensions in making the mandarin learning app success among our students. The success of the mandarin learning app is due to high student motivation, high cognitive development from the games plays and the games meet the student expectation. Meanwhile, the games are user friendly.

A set of questionnaires with 5 constructs (motivation, attitudes, cognitive development, interface design and expectation) with 24 items and an open-ended question on student’s open comments about learning using the games is designed as shown in Table III. We present the correlation plot and exploratory factor analysis for questionnaire items in all 5 constructs to answer derived on this goal.

TABLE III. THE DETAILED OF QUESTIONNAIRE WITH 5 CONSTRUCTS

Section	Item	Question
Student’s motivation	A1	I think this gaming activity gives me lots of benefits.
	A2	I prefer to answer questions using mobile games as compared to using books or paper.
	A3	I am very interested in using games for learning new Mandarin words in the future.
	A4	Digital learning games do not bring additional value to Mandarin language learning
	A5	I prefer to do exercises in games rather than quizzes during class.
	A6	The usage of computer games makes this subject more interesting.
Student’s attitudes with the games	B1	With the games, I can learn better by myself.
	B2	I can learn better according to my own pace and sequence.
	B3	It is more flexible for me to determine my own learning time.
	B4	It is more flexible for me to choose my learning pace.
	B5	The content of the games matches my subject syllabus.
	B6	The element of multiplayer in games motivates me to study Mandarin characters.
	B7	The games do not affect my motivation to learn new words.
Students’ cognitive development	C1	These computer games help me to think critically.
	C2	Solving the given problems is very interesting.
	C3	It is worth using games for learning in future.
	C4	Looking for the answer to questions given is an encouraging activity.
	C5	These games challenge my understanding of the subject.
Game interface	D1	Menus available in the games are easy to understand.
	D2	Navigations and interactions are easy to use.
	D3	Multimedia elements in the games are interesting.
	D4	I just need a very short time to know how the game is functioning.
	D5	The use of colour and design layout in the games are interesting.
Students expectation	E1	I wish I have more opportunities to learn other Mandarin course using this game approach.
	E2	I prefer using games to learn Mandarin as compared to traditional methods in the class.
	E3	I would like to learn all computer subjects using educational games.
	E4	I wish this game will be available online for easy access.

We carried out the analysis using RStudio, specifically by using package ‘psych’ [18]. We first reverse the Likert score for variables A4 and B7 as they are negatively phrased items (hereinafter, the terms “variable” and “item” are used interchangeably). As Likert-scale questionnaires yield ordinal data, we measure the association of the ordinal variables in terms of polychoric correlation and estimate the internal

consistency for scales using ordinal alpha. The coefficient ordinal alpha gives 0.94 indicating a very good level of reliability. Furthermore, this alpha value does not increase a lot when any of the items is deleted. Therefore, all the items as given in Table III are considered for further analysis.

We explore the strength of the relationship between all variables by constructing correlation plot with package ‘corrplot’ [19], as shown in Fig. 7. The correlation is positive (resp. negative) when one variable increases as the other increases (resp. decreases). We find that the two negatively phrased items (i.e. A4 and B7) as well as item D4 have negative correlation with all other variables, suggesting that item D4 probably should be reversed. However, we chose not to reverse item D4 scoring as it showed that our respondents understood the questions correctly.

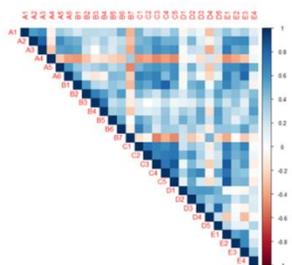


Fig. 7. Correlation Plot for all Likert-type Questionnaire Items.

We then employed multivariate statistical technique to further identify the possible latent relational structure among all the variables. Particularly, we performed exploratory factor analysis (EFA) to reexamine the grouping of variables and establish underlying dimensions that could explain the correlations, thereby allowing the formation and refinement of theory in the context of using mobile games to support Mandarin learning to a group of non-native learners. In other words, we tested the construct (i.e. factor) validity of the questionnaire with the factor analysis. The factors that explain the highest proportion of variance the variables share is expected to represent the underlying constructs [21].

To determine the suitability of our Likert-type questionnaire data for EFA, we perform Barlett’s test, and it returns p-value < 0.05 showing that the items correlate anyhow and EFA may be useful. Besides, the positive determinant of the correlation matrix suggests that EFA will probably run. We then determine the number of factors to retain by constructing the screen plot, as illustrated in Fig. 8. Together with parallel analysis, we take the number of factors as two for further analysis.

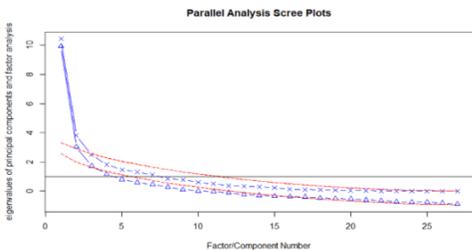


Fig. 8. The Screen Plot based on the Unreduced Correlation Matrix and Parallel Analysis.

We then carried out the exploratory factor analysis using maximum likelihood estimation procedure and its results are depicted in Fig. 9. Seventeen Likert-type items in the questionnaire make up the first factor and 9 items the second factor. Item E4 may be ignored as its load is lower than 0.3 for both factors. This implies that the measurement on students’ expectation is irrespective of whether the mobile game is available online or not. Also, from such a regrouping of variables, we could say that the correlations among all the variables might be categorized into two broad dimensions, in which the first dimension includes the students’ motivation, cognitive development and expectation whereas the second dimension mainly covers game interface. It is interesting to point out that variables related to students’ attitude with the mobile game are split into two dimensions, whereby items B1, B6 and B7 fall under first dimension while the remaining items B2 to B5 are in second dimension.

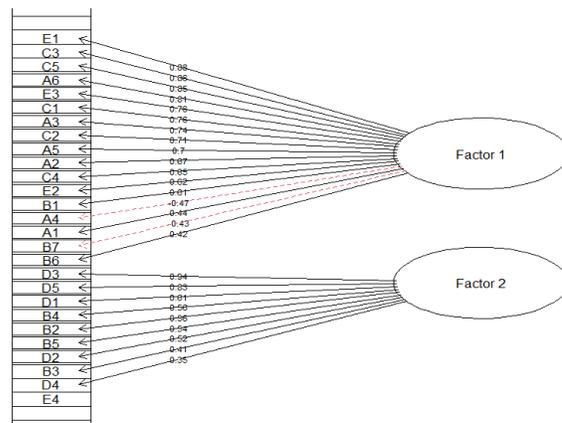


Fig. 9. Regrouping of Variables using Exploratory Factor Analysis.

V. CONCLUSION

This paper presents the adoption of GQM as an interdisciplinary tool to access the mobile mandarin app. With the GQM, we can summary the benefits of GQM as following. 1) Better structure of the experiments; 2) able to reach consensus among researchers from different disciplines; 3) able to analysis the dependencies among various experiments; 4) able to find hidden results. The GQM can structure the experiments and drive the communication among the researchers. It is also able to consolidate various experiments and preset the results in a more systematic manner as described in the previous section. In addition, the GQM can reduce the overlapping of method used, parameters and questions among different evaluation techniques that are used by the researchers. With GQM, we can discover the relationship across the experiments that are conducted by researchers. For example, the pre-test and post test results are able to continue study the cross culture learning among the students. The results can provide the answer to the question in achieving the goal of 'ensure cross culture learning'. On the other hand, the number of dimension and list of constructs can further be extended with factors like security and privacy to ensure the achievement of subgoals, 'ensure secure learning' and 'ensure privacy protection'. The contribution of this research is to demonstrate how GQM can be used as an interdisciplinary tool when assessing mobile learning application among researchers

from software engineering SE, language study, and mathematics. From the results, the mandarin learning app can serve as a complemented tool on conventional mandarin lesson. The post test results are more than pre- test results. Hence, the games can improve student results. They are two major dimensions in making the mandarin learning app success among our students. The success of the mandarin learning app is due to high student motivation, high cognitive development from the games plays and the games meet the student expectation. Meanwhile, the games are user friendly. In future, more works are required to investigate the adoption of others concept from agent models [22][23][24] into GQM. For example, to use HOMER questions in forming the questions is yet to explore.

ACKNOWLEDGMENT

The authors would like to thank UNIMAS to support on this research. This research is under UNIMAS Teaching and Learning grant, SoTL SoTL/FSKTM/2020(1)/002.

REFERENCES

- [1] Bokingito Jr, P. B., and Caparida, L. T., "Usability evaluation of a real-time water quality monitoring mobile application," *Procedia Computer Science*, 197, pp. 642-649, 2022.
- [2] Saleh, A., Ismail, R., and Fabil, N., "Evaluating usability for mobile application: a MAUEM approach," in *Proc. of the 2017 Int. Conf. on Software and e-Business*, pp. 71-77, Dec. 2017.
- [3] Kureerung, P., and Ramingwong, L., "Factors supporting user interface design of mobile government application," in *Proc. of the 2019 2nd Int. Conf. on Information Science and Systems*, pp. 115-119, Mar. 2019.
- [4] Gharaat, M., Sharbaf, M., Zamani, B., and Hamou-Lhadj, A., "ALBA: a model-driven framework for the automatic generation of android location-based apps," *Automated Software Engineering*, 28(1), pp. 1-45, 2021.
- [5] Hashim, N. L., and Isse, A. J., "Usability evaluation metrics of tourism mobile applications," *J. of Software Eng. and App.*, 12(7), pp. 267-277, 2019.
- [6] Noei, E., Zhang, F., Wang, S., and Zou, Y., "Towards prioritizing user-related issue reports of mobile applications," *Empirical Software Eng.*, 24(4), pp. 1964-1996, 2019.
- [7] Hussain, A. and Mohamed Omar, A., "Usability Evaluation Model for Mobile Visually Impaired Applications," *Int. Assoc. of Online Eng.*, 2020. [Online]. Available: <https://www.learntechlib.org/p/216427/> . Accessed on: Feb. 14, 2022.
- [8] Schobel, J., Pryss, R., Schlee, W., Probst, T., Gebhardt, D., Schickler, M., and Reichert, M., "Development of mobile data collection applications by domain experts: Experimental results from a usability study," in *Int. Conf. on Adv. Information Systems Eng.*, Springer, Cham, pp. 60-75, June 2019.
- [9] Tabora Mosquera, J. P., Arango-Lopez, J., Gutierrez Vela, F. L., Collazos, C., and Moreira, F., "Analyzing effectiveness and fun through metrics applied to pervasive gaming experiences," *Universal Access in the Information Society*, 20(3), pp. 545-554, 2021.
- [10] Khomyakov, I., Masyagin, S., and Succi, G., "Experience of Mixed Learning Strategies in Teaching Lean Software Development to Third Year Undergraduate Students," in *Int. Workshop on Frontiers in Software Eng. Education*, Springer, Cham, pp. 42-59, Nov. 2019.
- [11] Arifin, Y., Sastria, T. G., and Barlian, E., "User experience metric for augmented reality application: a review," *Procedia Computer Science*, 135, pp. 648-656, 2018.
- [12] Saleh, M. D. R. B., and Majrashi, K., "An enhanced usability model for mobile health application," *Int. Journal of Computer Science and Information Security (IJCSIS)*, 17(2), 2019.
- [13] Tahir, R., and Arif, F., "A measurement model based on usability metrics for mobile learning user interface for children," *The Int. Journal of E-Learning and Educational Technologies in the Digital Media*, 1(1), pp. 16-31, 2015.
- [14] Carvalho, R. M., de Castro Andrade, R. M., and de Oliveira, K. M., "AQUARIUM-A suite of software measures for HCI quality evaluation of ubiquitous mobile applications," *J. of Systems and Software*, 136, pp. 101-136, 2018.
- [15] Shiang, C. W., Halin, A. A., Lu, M., and CheeWhye, G., "Long Lamai Community ICT4D E-Commerce System Modelling: An Agent Oriented Role-Based Approach," *The Electronic J. of Information Systems in Developing Countries*, 75(1), pp. 1-22, 2016.
- [16] Alqahtani, A., Solaiman, E., Patel, P., Dustdar, S., and Ranjan, R., "Service level agreement specification for end-to-end IoT application ecosystems," *Software: Practice and Experience*, 49(12), pp. 1689-1711, 2019.
- [17] Ngadiman, N., Sulaiman, S., Idris, N., Mohamed, H., and Osman, S. M., "A Survey on Quality Factors in Designing Educational Applications for Active Learning," in *2019 IEEE 9th Int. Conf. on System Engineering and Technology (ICSET)*, IEEE, pp. 23-28, Oct. 2019.
- [18] Revelle, W., and Revelle, M. W., "Package 'psych'," *The Comprehensive R Archive Network*, pp. 338-338, 2015.
- [19] Wei, T., Simko, V., Levy, M., Xie, Y., Jin, Y., and Zemla, J., "Package 'corrplot'," *Statistician*, 56(316), 2017.
- [20] Barnett, A. G., Van Der Pols, J. C., and Dobson, A. J., "Regression to the mean: what it is and how to deal with it," *Int. J. of Epidemiology*, 34(1), pp. 215-220, 2015.
- [21] Taherdoost, H., Sahibuddin, S., and Jalaliyoon, N., "Exploratory factor analysis: concepts and theory," *Adv. in Applied and Pure Mathematics*, 27, WSEAS, pp. 375-382, 2014.
- [22] Cheah, W., ChinHong, P., Halim, A. A., "Agent-Oriented Requirement Engineering for Mobile Application Development," *International Journal of Interactive Mobile Technologies*, 11(6), 2017.
- [23] Zulkifli, S. F. B., Waishiang, C., Khairuddin, M. A. B., Jali, N. B., & Bujang, Y. R. B., "How to Model an Engaging Online Quiz? The Emotion Modeling Approach," *Journal of Telecommunications and Information Technology*, (1), pp. 54-63, 2022.
- [24] Ten LiBin, M., WaiShiang, C., Khairuddin, M. A. B., Mit, E., Erianda, A., "Agent-Oriented Modelling for Blockchain Application Development: Feasibility Study," *JOIV: International Journal on Informatics Visualization*, 5(3), pp. 248-255, 2021.