Liu, C.-C. et al. (Eds.) (2014). Proceedings of the 22nd International Conference on Computers in Education. Japan: Asia-Pacific Society for Computers in Education

Mobile-Assisted Learning Experimental Design: Current Deficiencies and Potential Improvements

Yao-Ting SUNG[a], Han-Yueh LEE[b*], Je-Ming YANG[b], Wei-Cheng CHEN[b], Hak-Ping TAM[c], Yi-Shian LEE[d]

[a]*Department of Educational Psychology and Counseling, National Taiwan Normal University,Taiwan*
[b]*Research Center for Psychological and Educational Testing, National Taiwan Normal University, Taiwan*
[c]*Graduate Institute of Science Education, National Taiwan Normal University, Taipei, Taiwan*
[d]*Aim for the Top University Project Office, National Taiwan Normal University, Taipei, Taiwan*
*ckshoupon@hotmail.com

Abstract: Due to the fact that mobile devices and educational software developed and expanded substantially in recent years, the importance of mobile-assisted learning may be raised. This study aimed to investigate the lack of the experimental design mobile-assisted leaning studies, and tried to make suggestions. Researchers collected all of the experimental studies on mobile-assisted learning published in ERIC and ISI from 2003 to 2013, which amounted to 216 studies. The four primary variables (research design, methods for initial equivalence in quasi-experiment, measuring tools and their reliability and validity, and sample size) were coded and preliminarily analyzed into frequency tables. Results of the study are as follows: (1) Researchers of 33.33% studies conducted the research by pre-experimental design or quasi-experimental design without equating, which were not rigorous sufficiently; (2) reliability and validity of outcome measures were not reported in 80% studies; and (3) sample sizes of approximate 30% studies were inadequate to draw an accurate statistical conclusion.

Keywords: Mobile-assisted learning, Design of experiments, Rigorous

1. Introduction

Along with the rapid development of mobile technology and software, mobile-assisted learning has already been widely adopted. In all mobile-assisted learning, a learner uses a mobile device (such as a smart phone, personal digital assistant, tablet PC, laptop, or other portable computer) to help them learn (Hwang, Tsai and Yang, 2008; Shih, Chu, Hwang and Kinshuk, 2011). There are several advantages of using a mobile device, such as the powerful calculating, high portability, wireless internet, instant communications, and context awareness. The advantages make mobile-devices become a tool with great potential for both traditional formal learning and informal outdoor learning. Although there were lots of experimental design studies exploring the effect of mobile-assisted learning, Cheung and Slavin (2013) warned that the results must be interpreted with caution because of serious methodological problems. They sorted out the common problems including the lack of a control group, limited evidence of initial equivalence between experimental and control groups, questionable outcome measures, small sample sizes, etc. Therefore, to examine the rigor of experimental design and improve the quality of educational experimental research will become increasingly important.

Valentine and Cooper (2008) proposed a research quality assessment scale named Study Design and Implementation Assessment Device (DIAD). DIAD includes four dimensions, which are internal validity, construct validity, external validity, and statistical conclusion validity. Internal validity refers to the validity of inferences about whether some intervention has caused an observed outcome. Construct validity refers to the extent of representative that intervention and outcome measures are supposed to be. External validity is the degree to which the results of a research can be generalized to other populations, settings, time, treatment variations, or outcomes. Statistical conclusion validity is the precision of outcome estimation. Based on the four dimensions proposed above, the purpose of this

study is to examine the current deficiencies and potential improvements in experimental design about mobile-assisted learning researches in last ten years.

2. Method

5.1 Data Sources, Search Strategy, and Search Results

This research employed electronic search to retrieve journal articles published since 2003 to 2013. The main databases were the Education Resources Information Center (ERIC) and the Social Sciences Citation Index (SSCI) database of the Institute of Science Index (ISI). Two sets of keywords were used: (1) mobile-device related keywords, including mobile, wireless, ubiquitous, wearable, portable, handheld, mobile phone, personal digital assistant (PDA), palmtop, pad, web pad, tablet PC, tablet computer, laptop, e-book, digital pen, pocket dictionary, and classroom response system; and (2) learning related keywords, including teaching, learning, training, and lecture. The two sets of keywords were combined to search the databases. The search yielded 4121 relevant literatures, and then reviewed by three researchers to assess their appropriateness for this study. Literatures were excluded if not using experimentation as research method. 216 articles were retained last for further analysis.

5.2 Variables Selection and Coding

To examine the studies' experimental design features, we chosen and coded four variables referred to the four dimensions of DIAD. The coded variables are as follows: research design (internal validity), methods for initial equivalence in quasi-experimental design (internal validity), measuring tools and their reliability and validity (construct validity), and sample size (external validity and statistical conclusion validity). Study coding was conducted by three researchers working independently. When disagreements occurred, the researchers inspected the studies in question together and reached a final agreement.

3. Result and Discussions

5.3 Internal validity

As shown in Table 1, we first found that true experimental studies are 35.65% of the total. This implies that only about a third of the studies are highly rigorous. Second, quasi-experimental studies took up the largest proportion (49.54%) of all research designs. Since quasi-experimental design lack the element of random assignment to experiment and control group, baseline comparability becomes an important issue. We further analyzed whether researchers used any of the method to ensure initial equivalence or not. There are 62.62% studies took two groups comparability into consideration by using adequate equating methods such as: (1) using t-test to confirm there are no significant difference between two groups in pre-test scores; (2) using gain scores (post-test – pre-test) as dependent variable to compare two groups; or (3) using ANCOVA or MANCOVA to exclude the difference between baseline ability of the two groups. However, there are still 37.38% studies didn't use any adequate equating procedures. Third, pre-experimental studies accounted for 14.82% of the total. This design cannot discriminate the treatment effect between intervention and participants' self-growth because of lacking a control group. To sum up, 72 of the 216 studies (33.33 %) were not rigorous sufficiently if we add pre-experimental studies to quasi-experimental studies that not using any equating procedures.

Table 1: Research design and methods for initial equivalence in quasi-experimental design.

Research design	Rigorous	*N* of articles	% of total	% of QEAs
1. Pre-experimental design	Not rig	32	14.82%	-
2. Quasi-experimental design	-	107	49.54%	100%
2.1 t-test of pre-test scores are *ns*	Rig	26	-	24.30%
2.2 Using gain scores as dependent variable	Rig	6	-	5.61%
2.3 Using ANCOVA or MANCOVA	Rig	35	-	32.71%
2.4 Not using any adequate equating method	Not rig	40	-	37.38%

3. True experimental design	Highly rig	77	35.65%	-

Note. N = number; QEAs = Quasi-experimental articles; rig = rigorous; *ns* = none significant.

3.1 Construct validity

According to the examination about the measuring tools and their respective reliability and validity (see Table 2) among the literatures, only 39 studies reported test reliability and validity both, which was only 18.06% of the total. Since decent test reliability and validity is the cornerstone of all rigorous research, approximate 80% of the research on this topic could make improvement.

Table 2: Measuring tools and their reliability and validity.

Measuring Tool / Reliability and validity	Number of articles	Percent of total
1.Self-design test/ unreported	99	45.83%
2.Self-design test/only reliability or validity	57	26.39%
3.Self-design test/ both reliability and validity	34	15.74%
4. Standardized test / unreported	15	6.94%
5. Standardized test / only reliability or validity	6	2.78%
6. Standardized test / both reliability and validity	5	2.32%

3.2 External validity and Statistical conclusion validity

Based on the description of various sample sizes (see Table 3), 73 studies had samples smaller than 50 people. According to Valentine and Cooper (2008), sample sizes must be at least 50 people or more in order to predict results fully and accurately. It means the results of 33.80% of the quantitative literature on mobile-assisted learning needs to be re-examined.

Table 3: Sample size.

Sample size	Number of articles	Percent of total
1. Below 50	73	33.80%
2. 50-99	71	32.87%
3.100-149	29	13.43%
4.150-199	11	5.09%
5. 200-249	8	3.70%
6. Above 250	24	11.11%

4. Conclusion and Future Work

According to the above, the standards of experiments on mobile-assisted learning need to be raised higher. Researchers could consider the 4 elements that DIAD proposed (internal validity, construct validity, external validity, and statistical conclusion validity) as a research criterion, and deal with them more rigorously in new studies. Furthermore, researchers could conduct a meta-analysis to compare and discuss the effects of various variables in the field.

References

Cheung, C. K., & Slavin, R. E. (2013). The effectiveness of educational technology applications for enhancing mathematics achievement in K-12 classrooms: A meta-analysis. *Educational Research Review 9*, 88-113.

Hwang, G. J., Tsai, C. C., & Yang, S. J. H. (2008). Criteria, strategies and research issues of context-aware ubiquitous learning. *Educational Technology & Society 11*(2), 81-91.

Shih, J. L., Chu, H. C., Hwang, G. J., & Kinshuk. (2011). An investigation of attitudes of students and teachers about participating in a context-aware ubiquitous learning activity. *British Journal of Educational Technology 42*(3), 373-394.

Valentine, J. C., & Cooper, H. (2008). A systematic and transparent approach for assessing the methodological quality of intervention effectiveness research: The study design and implementation assessment device (Study DIAD). *Psychological Methods, 13*(2), 130-149.