Effect of Using Assist Devices on Exam Completion Questions Among Thai College Students

Arthur E. Schneider

Lecturer, Department of Education, St. Theresa International College, Thailand 26120

Abstract

Action research was undertaken to begin to explore the possibility of improving second-language Thai college student performance on completion questions by using bolded and underscored words in test item stems, called 'assist devices'. This intervention was designed to focus student attention on key terms. Twenty-one students, in an intact class, were exposed to two course exams or testing trials, which included a total of 30 completion question, 15 of which were randomly selected for treatment with assist devices, and thus used as the experimental condition. The 15 remaining, randomly selected questions, were not treated and thus used as the control condition. The results of the first trial were invalidated due to an unforeseen error in the online exam. Thus, only the results of the second trial were analyzed using One-way Repeated Measures ANOVA, which indicated a significant treatment effect. While the results were encouraging, any substantive conclusions on the use of assist devices on completion test items with second-language Thai college students will have to wait for further longitudinal testing to establish a possible pattern of correlation. While a strong correlation was evident, a valid cause and effect relationship could not be established due to the measurement error introduced by unknown item discrimination data, which was not explored in this pilot testing phase. The use of randomly selected, treated and untreated questions in the same trial, however, was thought to be an innovative design strategy for avoiding the need of random assignment of subjects into experimental and control groups, thereby making a valuable contribution to the action research paradigm.

Keywords: Learner-centered assessment, Completion questions, Thai college students, Second-language students

1. Introduction

Should not learner-centered assessment be included in learner-centered education? In the past, learner-centered assessment has taken different forms (see Duncan & Buskirk-Cohen, 2011; Wright & Graham, 1975), which attempted to take into account the differential aptitude levels of students. The learner-centered assessment approach taken in this action research (Anderman, 2009) pilot study encompassed using what were termed 'assist devices' in completion test items, to help students focus on key words, thus potentially facilitating a better understanding and greater accuracy in responding to such treated test questions. This is consistent with the supposition that devices used to draw attention generally result in a kind of self-teaching or active learning (Michael, 2006), and better performance on retention, where the highlighted items help readers to discriminate between important material and background material (Fowler & Barker, 1974).

2. Literature Review

No directly related studies were found regarding the study design used, specifically the use of assessment intervention on completion test questions and exposure of subjects to both control and experimental conditions in the same trial. However, theory support for various aspects of this research does come from the work of a number of sources.

A completion test question format, with one- or two- word answers, was used because it is an effective overt measure or observable manifestation of what students actually know (Wright & Graham, 1975), with the advantage of essentially having bimodal scoring (i.e., correct or incorrect), as answers are very brief. Multiple-choice, true-false, or matching questions, in contrast, do not require production of textual answers and are thus more subject to guessing, which vitiated their use in this study. Longer completion or essay questions, answers for which may be partially correct (i.e., not bimodal), were therefore also excluded from the study focus.

For the purposes of this pilot study, the argument as to whether completion questions should involve some level of critical thinking was not a factor. More precisely, the completion questions used in the study were part of a larger examination that included short-answer and essay questions that were designed to elicit critical thinking, as well as true-false and multiple-choice questions. Different types of test questions, per standard test design protocol, were used to allow students to have opportunities to express their differential test taking abilities (Brown, Race, & Smith, 1996).

It may be argued that providing assist devices on completion items unfairly helps some students to achieve a higher score compared to other students who do not need such scaffolding (Bruner, 1960). However, it must be stressed that 'achievement' and 'capability' (aptitude) are not the same thing (Shavelson, et al., 2002). That is, the ability of a student to benefit from an assist device on a completion question is a valid differential

indicator of capability because it is a better measure of 'whole student' aptitude (OECD, 2008). Indeed, this is the basic goal behind the widely supported educational scaffolding structure called Formative Assessment, in which teachers use a host of needs-based learning support mechanisms to help students develop to their full potential (OECD, 2008).

Furthermore, regarding the Summative Assessment used in this study, being able to identify those who cannot benefit from test item intervention is a valuable tool for teachers, as it helps them to more precisely identify and target different levels of aptitude in a given group of students. Students evincing such differential performance levels depending on external guidance is consistent with Vygotsky's (1978) Zone of Proximal Development, which is the difference in student performance with teacher intervention versus without.

However, it is argued that the only time that assessment assist devices would be helpful to students is when they are on the precipice of getting the correct answer. Thus, the distance between those who know the correct answer without assistance and those who discover the correct answer with assistance is necessarily small but, at the same time, a more precise measure of student capability.

Another argument for using assist devices in second-language classroom assessment is that of motivation. It is well known that second-language learners tend to be significantly motivated in learning situations when hints or clues are provided that may facilitate their getting correct answers (Krashen, 2013).

General theoretical support for the study design also comes from Constructivist learning theory (Piaget, 2001; Bruner, 1960), which would logically argue that assist devices on exams facilitate the interaction of the test taker with key exam elements so that they may better construct their understanding of test questions. Arguably then, use of such devices should actually improve test validity.

It may be more broadly argued that many students are capable of achieving at an overall good level of knowledge in a given subject, while a small percentage of students would be capable what might be called professional level knowledge. This marking system is used in the UK universities, where scoring between 70% and 90% will receive an A and is considered 'first class knowledge', but scoring above 90% is considered professional level knowledge. This system makes sense at all educational levels because some students actually do exhibit extraordinary knowledge on summative and formative assessments throughout their schooling. It also makes sense since students who work hard in math, science, English, etc., should be rewarded with a good mark, even though they will not become professionals in those fields. With scaffolding in all areas of education, including testing (i.e., 360 degree scaffolding), more students may become capable of achieving a good level of knowledge about different subjects, but those few who score extraordinarily well certainly will not be harmed by such support. It is also arguably possible to usher a few students into an upper level of achievement with comprehensive scaffolding. Some students, for example, have confidence problems (Bruner, 1963) or readiness issues (Tomlinson et al., 2003) that hold them back from their true potential, and this is where 360 degree scaffolding may make a difference.

Thus, given the above, it is logical to hypothesize that students' understanding of and performance on completion test questions would be facilitated by using assist devices in item stems. Action research was used to verify this hypothesis among second-language Thai students in an Education course, entitled Innovation in Education and Technology, in a small college in Thailand.

3. Method

The purpose of doing this research project is consistent with what is termed Canonical Action Research (Järvinen, 2012), which attempts to improve teaching and learning by using the basic tenets of the scientific method (question, hypothesis, intervention, interpretation of results, and innovation) (Rothchild, 2006), but which is bounded by the realities of intact classroom research. Thus, for the current study, an action research paradigm was developed to explore the effects of using assist devices on completion question performance among Thai college students in an intact classroom.

The term 'assist device' herein is defined as using bolded and underscored words to focus students' attention on key terms in test item stems. The purpose of providing such markers was to improve student understanding of completion questions and therefore improve performance. Questions that contained assist devices were termed 'treated questions' and those without such devices, 'untreated' questions. The study sample of 21 students was exposed to a randomly selected set of 15 treated questions and 15 untreated questions within the same exam or first trial. This pattern was repeated in a second exam (i.e., second trial), about six weeks later. No questions were shared between the two exams. Two phases of treatment using the same students were used in order to check for a repeated correlation pattern between the use of assist devices and greater accuracy in test item responses.

In a true experimental design, subjects would be randomly assigned to either an experimental group or a control group. In a typical educational setting, there are three obstacles to such a design. One, there may be insufficient numbers of students available to make the study statistically valid. Two, random assignment is not possible with intact classrooms. Three, some scholars view exposing some students to a potentially beneficial

educational intervention, while others are not as being unethical. A reasonable compromise for an action research design, in this case, is to avoid the need for random assignment into experimental and control groups by exposing all subjects to both the experimental and control conditions within the same trial.

Another part of standard study design would be to test items for item discrimination. However, the current action research study was meant to mirror the limitations of authentic teaching situations, where there is no time to pre-test questions and cull them using item discrimination statistics. In this is a pilot study, the aim of designing questions with good item discrimination was targeted by using existing guides for test question design (see Clay, 2001; Wills, 2001; Center for Innovation, n.d.), tying test items directly to content and terminology covered in the course (i.e., testing recall of course materials rather than critical thinking), attempting to create all completion questions with the same level of challenge (i.e., 'intermediate', based on the experimenter's extensive experience in test item design), and by randomly assigning questions to the treated pool or the untreated pool. Random assignment should randomly distribute questions that have better item discrimination with those who weaker item discrimination, so that the treated pool and the untreated pool are comparable. Developing completion question design based on these criteria should strengthen average item discrimination.

4. Results and Discussion

The results of the first trial had to be discarded due to an unforeseen technical error in the online test. Therefore, only the results of the second trial are reported below using analyses from a one-way, repeated measures ANOVA, which was the closest fit to the study design, even though the experimental and control conditions were administered in one occasion rather than two.

Therefore a one-way, within subjects ANOVA was conducted to determine the effect of completion test item treatment on the average number of correct answers students achieved on treated versus untreated completion questions.

Descriptive statistics for trial 2 can be seen in Table 1 below. The control condition (T2CONTROL_FINAL) mean indicates the number of questions that 21 students answered correctly, on average, from the 15 randomly chosen and treated completion questions on the exam. The experimental condition (T2EXP_FINAL) mean indicates the number of questions the same 21 students answered correctly, on average, from the 15 randomly chosen, but untreated, completion questions. It is clear in Table 1 that the experimental condition (M = 12.00, SD = 1.22) generated a higher mean number of correct answers, as compared to the control condition (M = 9.86, SD = 1.06)

Table 1. Descriptive statistics

	Mean	Std. Deviation	N
T2CONTROL__FINAL	9.86	1.06	21
T2EXP__FINAL	12.00	1.22	21

Table 2 indicates that the treatment effect was significant, $F(1, 20) = 63.08$, $p < .01$, suggesting a potential cause and effect relationship between question treatment and student performance in the limited context of this study.

Table 2. Treatment effects

Source	Treatment	Type III Sum of Squares	df	Mean Square	F	Sig.
Treatment	Linear	48.214	1	48.214	63.084	.000
Error(Treatment)	Linear	15.286	20	.764		

5. Conclusion

While the results are encouraging, any substantive conclusions as to the use of assist devices on completion question with second-language Thai college students will have to wait for further longitudinal testing to establish a possible pattern of correlation. While a strong correlation is evident, a cause and effect relationship could not be established due to the measurement error introduced by unknown item discrimination data, which was not explored in this pilot testing phase. The use of randomly selected, treated and untreated questions in the same trial, however, was thought to be an innovative design strategy for avoiding the need of random assignment of subjects into experimental and control groups, which is a major limitation in the action research paradigm. Thus, the action research goal of generating information that is useful to critically thinking about learner-centered assessment was met.

6. References
Anderman, E. (2009). *Research methods: An overview*. Farmington Hills, MI: The Gale Group. Retrieved from http://www.education.com/reference/article/research-methods-an-overview/#B

Brown, S., Race, P., & Smith, B. (1996). *500 Tips on assessment*. London (UK): Kogan Page. Retrieved from: https://www.google.co.th/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=0ah UKEwjKo9vvxITRAhUIro8KHW4cAE8QFggYMAA&url=http%3A%2F%2Fold.medu.ir%2Fportal% 2FFile%2FShowFile.aspx%3FID%3Da7b29248-da20-43cd-9df3-8c622fef402d&usg=AFQjCNGNA1cgEHFKR_IzhnLCvJaPdHYErA

Bruner, J. S. (1960). *The process of education*. Cambridge (MA): Harvard University Press.

Bruner, J. S. (1963, May). Needed: A theory of instruction. *Educational Leadership 20*(8): 523-532. Retrieved from: http://www.ascd.org/ASCD/pdf/journals/ed_lead/el_196305_bruner.pdf

Center for Innovation in Teaching & Learning. (n.d.). *Improving your test questions*. Retrieved from: https://citl.illinois.edu/citl-101/measurement-evaluation/exam-scoring/improving-your-test-questions#completion

Clay, B. (2001). *Is this a trick question? A short guide to writing effective test questions*. Kansas (US): Kansas Curriculum Center. Retrieved from: https://www.k-state.edu/ksde/alp/resources/Handout-Module6.pdf

Duncan, T., & Buskirk-Cohen, A. A. (2011). Exploring learner-centered assessment: A cross-disciplinary approach. *International Journal of Teaching and Learning in Higher Education, 23*(2). 246-259. Retrieved from: http://files.eric.ed.gov/fulltext/EJ946150.pdf

Fowler, R. L., & Barker, A. S. (1974). Effectiveness of highlighting for retention of text material. *Journal of Applied Psychology, 59*(3), 358-364. Retrieved from: https://www.researchgate.net/publication/232539495_Effectiveness_of_highlighting_for_retention_of_t ext_material

Järvinen, P. (2012). *On boundaries between field experiment, action research and design research*. Tampere, Finland: University of Tampere, School of Information Sciences. Retrieved from: http://www.uta.fi/sis/reports/index/R14_2012.pdf

Krashen, S. (2013) *Second language acquisition: Theory, application, and some conjecture*. Cambridge (UK): Cambridge University Press.

Michael, J. (2006, Dec). Where's the evidence that active learning works? *Advances in Physiology Education, 30*(4). 159-167. Retrieved from: http://advan.physiology.org/content/30/4/159.long

OECD. (2008). Assessment for learning: Formative assessment, presented at OECD/CERI International Conference Learning in the 21st Century: "Research, Innovation and Policy", Paris, 2008. Paris: OECD. Retrieved from: http://www.oecd.org/site/educeri21st/40600533.pdf

Piaget, J. (2001). *The psychology of intelligence*. Abingdon-on-Thames (UK): Routledge.

Rothchild, I. (2006). *Induction, deduction, and the scientific method: An eclectic overview of the practice of science*. Madison, WI: Society for the Study of Reproduction, Inc. Retrieved from: *http://www.ssr.org/Documents/2006-01-04Induction2.pdf*.

Shavelson, R. J., Roeser, R. W., Kupermintz, H., Lau, S., Ayala, C., Haydel, A., …Quihuis, G. (2002). Richard E. Snow remaking of the concept of aptitude and multidimensional test validity: Introduction to the special issue. *Educational Assessment, 8*(2). 77-99. Retrieved from: http://web.stanford.edu/dept/SUSE/SEAL/Reports_Papers/Richard%20E%20Snow's%20Remaking%20 of%20the%20Concept%20of%20Aptitude%20and%20Mu.pdf

Tomlinson, C. S., Brighton, C., Hertberg, H., Callahan, C. M., Moon, T. R., Brimijoin, K., … Reynolds, T. (2003). Differentiating instruction in response to student readiness, interest, and learning profile in academically diverse classrooms: A Review of literature. *Journal for the Education of the Gifted, 27*(2/3), p. 119–145. Retrieved from: http://files.eric.ed.gov/fulltext/EJ787917.pdf

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Wills, S. (2001). *Designing effective test questions*. Masterminds, LLC. Retrieved from: http://www.calhoun.k12.al.us/makes%20sense/Adobe%20Reader/DO%20NOT%20OPEN%20program %20files/Assessment/designing%20test%20questions.pdf

Wright, B. D., & Graham, A. D. (1975). *Best test design and self-tailored testing*. Chicago (US): University of Chicago Department of Education. Retrieved from: http://mmm1406.sanjose14-verio.com/rascho/memo19.pdf